Predictive Modelling of Global Solar Radiation with Artificial Intelligence Approaches using MODIS Satellites and Atmospheric Reanalysis Data for Australia

A thesis submitted by

Sujan Ghimire

B.E (Mechanical Engineering)

M.E (Renewable Energy Engineering)

For the award of

Doctor of Philosophy

2019

Dedication

Dedicated to the memory of my mother, Mrs. Anju Ghimire who always believed in my ability to be successful in the academic arena. You are gone but your belief in me has made this journey possible.

Abstract

Global solar radiation (*GSR*) prediction is a prerequisite task for agricultural management and agronomic decisions, including photovoltaic (PV) power generation, biofuel exploration and several other bio-physical applications. Since short-term variabilities in the *GSR* incorporate stochastic and intermittent behaviours (such as periodic fluctuations, jumps and trends) due to the dynamicity of atmospheric variables, *GSR* predictions, as required for solar energy generation, is a challenging endeavour to satisfactorily predict the solar generated electricity in a PV system. Additionally, the solar radiation data, as required for solar energy monitoring purposes, are not available in all geographic locations due to the absence of meteorological stations and this is especially true for remote and regional solar powered sites. To surmount these challenges, the universally (and freely available) atmospheric gridded datasets (*e.g.*, reanalysis and satellite variables) integrated into solar radiation predictive models to generate reliable *GSR* predictions can be considered as a viable medium for future solar energy exploration, utilisation and management. Hence, this doctoral thesis aims to design and evaluate novel Artificial Intelligence (AI; Machine Learning and Deep Learning) based predictive models for *GSR* predictions, using the European Centre for Medium Range Weather Forecasting (ECMWF) Interim-ERA reanalysis and Moderate Resolution Imaging Spectroradiometer (MODIS) Satellite variables enriched with ground-based weather station datasets for the prediction of both long-term (*i.e.*, monthly averaged daily) as well as the short-term (*i.e.*, daily and half-hourly) *GSR*. The focus of the study region is Queensland, the sunshine state, as well as a number of major solar cities in Australia

where solar energy utilisation is actively being promoted by the Australian State and Federal Government agencies.

Firstly, the Artificial Neural Networks (ANN), a widely used Machine Learning model is implemented to predict daily *GSR* at five different cities in Australia using ECMWF Reanalysis fields obtained from the European Centre for Medium Range Weather Forecasting repository. Secondly, the Self-Adaptive Differential Evolutionary Extreme Learning Machine (*i.e.*, SaDE-ELM) is also proposed for monthly averaged daily *GSR* prediction trained with ECMWF reanalysis and MODIS satellite data from the Moderate Resolution Imaging Spectroradiometer. Thirdly, a three-phase Support Vector Regression (SVR; Machine Learning) model is developed to predict monthly averaged daily *GSR* prediction where the MODIS data are used to train and evaluate the model and the Particle Swarm Algorithm (PSO) is used as an input selection algorithm. The PSO selected inputs are further transformed into wavelet subseries via non-decimated Discrete Wavelet Transform to unveil the embedded features leading to a hybrid PSO-W-SVR model, seen to outperform the comparative hybrid models. Fourthly, to improve the accuracy of conventional techniques adopted for *GSR* prediction, Deep Learning (DL) approach based on Deep Belief Network (DBN) and Deep Neural Network (DNN) algorithms are developed to predict the monthly averaged daily *GSR* prediction using MODIS-based dataset. Finally, the Convolutional Neural Network (CNN) integrated with a Long Short-Term Memory Network (LSTM) model is used to construct a hybrid CLSTM model which is tested to predict the half-hourly *GSR* values over multiple time-step horizons (*i.e.*, 1-Day, 1-Week, 2-Week, and 1-Month periods). Here, several statistical, Machine Learning and Deep Learning models are adopted to benchmark the proposed DNN and CLSTM models against conventional models (ANN, SaDE-ELM, SVR, DBN).

In this doctoral research thesis, a Global Sensitivity Analysis method that attempts to utilise the Gaussian Emulation Machine (GEM-SA) algorithm is employed for a sensitivity analysis of the model predictors. Sensitivity analysis of selected predictors ascertains that the variables: aerosol, cloud, and water vapour parameters used as input parameters for *GSR* prediction play a significant role and the most important predictors are seen to vary with the geographic location of the tested study site. A suite of alternative models are also developed to evaluate the input datasets classified into El Niño, La Niña and the positive and negative phases of the Indian Ocean Dipole moment. This considers the impact of synoptic-scale climate phenomenon on long-term *GSR* predictions.

A seasonal analysis of models applied at the tested study sites showed that proposed predictive models are an ideal tool over several other comparative models used for *GSR* prediction. This study also ascertains that an Artificial Intelligence based predictive model integrated with ECMWF reanalysis and MODIS satellite data incorporating physical interactions of the *GSR* (and its variability) with the other important atmospheric variables can be considered to be an efficient method to predict *GSR*. In terms of their practical use, the models developed can be used to assist with solar energy modelling and monitoring in solar-rich sites that have diverse climatic conditions, to further support cleaner energy utilization.

The outcomes of this doctoral research program are expected to lead to new applications of Artificial Intelligence based predictive tools for *GSR* prediction, as these tools are able to capture the non-linear relationships between the predictor and the target variable (*GSR*). The Artificial Intelligence models can therefore assist climate adaptation and energy policymakers to devise new energy management devices not only for Australia but also globally, to enable optimal management of solar energy resources and promote renewable energy to combat current issues of climate change. Additionally, the proposed predictive models may also be applied to other renewable energy areas such as wind, drought, streamflow, flood and electricity demand for prediction.

Certification of Thesis

This thesis is the work of ***Sujan Ghimire*** except where otherwise acknowledged, with the majority of the authorship of the paper presented as a Thesis by Publication undertaken by the doctoral research student. The work is original and has not been previously been submitted for any other award, except where acknowledged.

Sujan Ghimire — *15 May 2019*

PhD Candidate — Date

Endorsement

Dr Ravinesh C Deo — *29 June 2019*

Principal Supervisor — Date

Dr Nawin Raj — *29 June 2019*

Associate Supervisor — Date

Dr Nathan J Downs — *29 June 2019*

Associate Supervisor — Date

Professor Jianchun Mi — *29 June 2019*

External Supervisor — Date

Student and supervisors' signatures of endorsement are held at USQ.

Statement of Contributions

The articles produced from this doctoral research thesis were a joint contribution of the student and the supervisors. The details of the scientific contribution of each author in the respective journal publications and book Chapter are provided as follows.

Article I: Chapter 3

Sujan Ghimire, Ravinesh C. Deo, Nathan J. Downs, and Nawin Raj. "Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cites of Queensland Australia". *Journal of Cleaner Production*, 216 (2019):288-310. [***Impact Factor: 6.680, SNIP: 2.308, Scopus Ranked Q1,96th percentile in Renewable Energy, Sustainability and the Environment].***
DOI: https://doi.org/10.1016/j.jclepro.2019.01.158

Author	Task Performed	Contribution
Sujan Ghimire (PhD Candidate)	Literature review, method development, programming, data analysis, preparation of figures and tables, compilation, writing, and revision of manuscript.	70%
Ravinesh C. Deo (Principal Supervisor)	Supervised and assisted in model concepts, provided detailed comments on the manuscript, edited and guided to prepare for submission.	20%
Nathan J. Downs (Associate Supervisor)	Comment on draft manuscript, detail grammar check and provide comments on language used.	5%
Nawin Raj (Associate Supervisor)	Proofreading and co-authorship of the manuscript.	5%

Article II: Chapter 4

Sujan Ghimire, Ravinesh C. Deo, Nathan J. Downs, and Nawin Raj. "Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed MODIS satellite and Reanalysis atmospheric products in solar-rich cities". *Remote Sensing of Environment,* 212 (2018): 176-198. [***Impact Factor: 8.100, SNIP: 2.961, Scopus Ranked Q1, 99th percentile in Earth and Planetary Sciences].***

DOI: https://doi.org/10.1016/j.rse.2018.05.003.

This paper has been awarded USQ PhD Publications Excellence Award 2019 (third category).

Author	Task Performed	Contribution
Sujan Ghimire (PhD Candidate)	Literature review, method development, programming, data analysis, preparation of figures and tables, compilation, writing, and revision of manuscript.	70%
Ravinesh C. Deo (Principal Supervisor)	Supervised And assisted in model concepts, provided detailed comments on the manuscript, edited and guided to prepare for submission.	20%
Nathan J. Downs (Associate Supervisor)	Comment on draft manuscript, detail grammar check and provide comments on language used.	5%
Nawin Raj (Associate Supervisor)	Comments on draft manuscript.	5%

Article III: Chapter 5

Sujan Ghimire, Ravinesh C. Deo, Jianchun Mi, and Nawin Raj." Wavelet-based 3-phase hybrid SVR model trained with particle swarm optimization and maximum overlap discrete wavelet transform for global solar radiation prediction with satellite-derived predictors". *Renewable & Sustainable Energy*

Reviews 113 (2019):109247. [***Impact Factor: 10.46, SNIP: 3.69, Scopus Ranked Q1,96th percentile in Renewable Energy, Sustainability and the Environment].***

DOI: https://doi.org/10.1016/j.rser.2019.109247

Author	Task Performed	Contribution
Sujan Ghimire (PhD Candidate)	Literature review, method development, programming, data analysis, preparation of figures and tables, compilation, writing, and revision of manuscript.	70%
Ravinesh C. Deo (Principal Supervisor)	Supervised and assisted in model concepts, provided detailed comments on the manuscript, edited and guided to prepare for submission.	20%
Jianchun Mi (External Supervisor)	Comment on draft manuscript, detail grammar check and provide comments on language used.	5%
Nawin Raj (Associate Supervisor)	Comment on draft manuscript, detail grammar check and provide comments on language used.	5%

Article IV: Chapter 6

Sujan Ghimire, Ravinesh C. Deo, Jianchun Mi, and Nawin Raj. "Deep learning neural network trained with MODIS satellite-derived predictors for long-term global solar radiation prediction". *Energies* . [***Impact Factor: 2.670, SNIP: 1.340, Scopus Ranked Q1, 84th percentile in Energy Engineering and Power Technology].***

DOI: https://doi.org/10.3390/en12122407

Author	Task Performed	Contribution
Sujan Ghimire (PhD Candidate)	Literature review, method development, programming, data analysis, preparation of figures and tables, compilation, writing, and revision of manuscript.	70%

Ravinesh C. Deo (Principal Supervisor)	Supervised and assisted in model concepts, provided detailed comments on the manuscript, edited and guided to prepare for submission.	20%
Jianchun Mi (External Supervisor)	Comment on draft manuscript, detail grammar check and provide comments on language used.	5%
Nawin Raj (Associate Supervisor)	Comment on draft manuscript, detail grammar check and provide comments on language used.	5%

Article V: Chapter 7

Sujan Ghimire, Ravinesh C. Deo, Jianchun Mi, and Nawin Raj." Deep Solar Radiation Forecasts with Hybrid Convolutional Neural Network integrated with Long Short-term Memory Network Algorithms". *Applied Energy,* 253 (2019):113541 [***Impact Factor: 8.57, SNIP: 2.62, Scopus Ranked Q1, 99th percentile in Energy].***

Url: https://doi.org/10.1016/j.apenergy.2019.113541

Author	Task Performed	Contribution
Sujan Ghimire (PhD Candidate)	Literature review, method development, programming, data analysis, preparation of figures and tables, compilation, writing, and revision of manuscript.	70%
Ravinesh C. Deo (Principal Supervisor)	Supervised and assisted in model concepts, provided detailed comments on the manuscript, edited and prepared for submission.	20%
Jianchun Mi (Associate Supervisor)	Comment on draft manuscript, detail grammar check and provide comments on language used.	5%
Nawin Raj (Associate Supervisor)	Comment on draft manuscript, detail grammar check and provide comments on language used.	5%

Book Chapter:

Sujan Ghimire, Ravinesh C. Deo, Nathan J. Downs, and Nawin Raj." Optimization of windspeed prediction using an artificial neural network compared with a genetic programming model". In: *Handbook of research on predictive modeling and optimization methods in science and engineering*. Advances in Computational Intelligence and Robotics (ACIR) Book Series. IGI Publishing (IGI Global), Hershey, United States, pp. 328-359. ISBN 9781522547662, Publication Date: January 2018

Url: https://www.igi-global.com/book/handbook-research-predictive-modeling-optimization/185480

Author	Task Performed	Contribution
Sujan Ghimire (PhD Candidate)	Literature review, method development, programming, data analysis, preparation of figures and tables, compilation, writing, and revision of manuscript.	75%
Ravinesh C. Deo (Principal Supervisor)	Supervised and assisted in model concepts, provided detailed comments on the manuscript, edited and guided to prepare for submission.	15%
Nathan J. Downs (Associate Supervisor)	Comment on draft manuscript, detail grammar check and provide comments on language used.	5%
Nawin Raj (Associate Supervisor)	Comment on draft manuscript, detail grammar check and provide comments on language used.	5%

Acknowledgements

I would like to thank all the people who made this thesis possible.

Sincere gratitude and appreciation towards my Principal Supervisor, Dr Ravinesh Deo for giving me the opportunity to find my ways towards a deeper understanding of Artificial Intelligence. His advice has been the reference that has allowed both to explore different fields of knowledge related to Machine Learning and to complete this doctoral research thesis. Obviously, several sentences are not enough to acknowledge his contribution to my achievement.

I would like to convey my appreciation to the Associate Supervisors, Dr Nawin Raj and Dr Nathan J. Downs and including External Supervisor Professor Jianchun Mi for providing further editorial support.

My appreciation also goes to the members of the "*Advanced Data Analytics*: Environmental Modelling & Simulation Research Group" led by the Principal Supervisor for contributions and my colleagues in sharing ideas.

A very special gratitude goes to the Research Training Scheme (RTS) provided by the Australian Government to the University of Southern Queensland for funding the study. I thank the University PhD Publication Excellence Award 2019 for a paper in '*Remote Sensing of Environment*" originating from this doctoral Thesis.

I would like to take this opportunity to thank all the organizations that provided freely accessible data including: the Scientific Information for Land Owners (SILO), Australian Bureau of Meteorology (BOM), the Goddard Earth Sciences Data and Information Services Center (GES DISC) and the Interim ERA European Centre for Medium-Range Weather Forecasting Reanalysis (ECMWF reanalysis).

I would especially like to thank my family. My wife, Anisha has been extremely supportive of me throughout this entire process and has made countless sacrifices to help me get to this point. My daughters Shreeyani and Reeyani, you are my inspiration to achieve greatness. Without you, I would not be where I am today. You all three have made me stronger, better, and more fulfilled than I could have ever imagined. I love you to the moon and back.

Next, I would like to thank my father Mr. Kedar Parsad Ghimire, father in law, Mr. Surendra Pokhrel, mother in law, Mrs. Anupa Pokhrel and grandmother in law, Mrs. Nirmala Koirala for all the sacrifices they have made to support my studies.

I am grateful to my sibling (Mr. Sanjeev Ghimire and Mrs. Sujeeta Sapkota) and Aunt Gauri and Muna, who have provided me through moral and emotional support in my life. I am also grateful to my other family members and colleagues who have supported me along the way.

I want to acknowledge the work of so many good teachers, from whom I have learned so much. Additionally, I would like to remember all the people that I have met during this time, and that in some way have contributed to this thesis.

Finally, I am grateful to have had the privilege of attending the prestigious University of Southern Queensland. This experience has afforded me the opportunity to work with some of the best and brightest, and the resources for me to achieve immense success. Thank you for this opportunity.

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Acronyms

ACO	Ant Colony Optimization
AIRS	Atmospheric Infrared Sounder
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
APB	Absolute Percentage Bias
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
BOM	Bureau of Meteorology

CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
CNN	Convolutional Neural Networks
CPU	Central Processing Unit
CRO	Coral Reef Optimization
CSIRO	Commonwealth Scientific and Industrial Research Organization
DBN	Deep Belief Network
DL	Deep Learning
DNN	Deep Neural Network
DSITIA Arts	Department of Science, Information Technology, Innovation and the
DT	Decision Tree
DWT	Discrete Wavelet Transformation
E_1/LM	Legates & McCabe's Index
ECMWF	European Centre for Medium-range Weather Forecasting
EEMD	Ensemble Empirical Mode Decomposition
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
EMS	Energy Management System
E_{NS}	Nash–Sutcliffe Efficiency
ENSO	El Nino-Southern Oscillation
EWT	Empirical Wavelet Transformation
EXH	Exhaustive Search
FE	Forecasting Error
FFA	Firefly Algorithm

FS/ IS	Feature Selection / Input Selection
fsrnca	Feature Selection for Regression based on Neighborhood Component
GA	Genetic Algorithm
GBM	Gradient Boosting Machine
GEM-SA	Gaussian Emulation Machine for Sensitivity Analysis
Giovanni	Goddard Online Interactive Visualization and Analysis Infrastructure
GMDH	Group Method of Data Handling
GP	Genetic Programming
GPML	Gaussian Process Machine Learning
GRU	Gated Recurrent Unit
GSR/I_{rad}	Global Solar Radiation
GSR_{obs}	Observed *GSR*
GSR_{pred}	Predicted *GSR*
IEA	International Energy Agency
IMF	Intrinsic Mode Function
IOD	Indian Ocean Dipole
IPCC	International Panel for Climate Change
KGE	Kling-Gupta Efficiency
L2	Lasso regularization
LM	Legates-McCabe's Index
LS-SVR	Least-Square SVR
LST	Land Surface Temperature
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error

MAPE	Mean Absolute Percentage Error
MARS	Multivariate Adaptive Regression Splines
MBE	Mean Bias error
MDA	Mean Decrease Accuracy
MDB	Murray-Darling Basin
MERRA	Modern Era Retrospective-analysis for Research and Applications
MIR	Mutual Information Regression
ML	Machine Learning
MLP	Multi-Layer Perceptron
MLR	Multi-Linear Regression
MODIS	Moderate Resolution Imaging Spectroradiometer
MODWT	Maximum Overlap Discrete Wavelet Transformation
MPMR	Minimax Probability Machine Regression
MRA	Multi-Resolution Analysis
MSE	Mean Squared Error
N	Length of dataset
NCA	Neighbourhood Component Analysis
NSGA	Nondominated Sorting Genetic Algorithm
NSW	New South Wales
OLS	Orthogonal Least Squares
OMI	Ozone Measuring Instrument
PACF	Partial Auto-Correlation Function
PE	Prediction Error
PSO	Particle Swarm Optimization

PV	Solar Photovoltaic
QLD	Queensland
r	Pearson's Correlation coefficient
RBM	Restricted Boltzmann Machine
r_{cross}	Cross-Correlation Function
Relieff	Relieff Algorithm
ReLU	Rectified Linear Unit
RES	Renewable Energy Sources
RF	Random Forest
RFER	Recursive Feature Elimination
RMAE	Relative MAE
RMSE	Root-Mean-Square-Error
RNN	Recurrent Neural Network
RRMSE	Relative Root-Mean-Square Error
SA	Simulated Annealing
SBR	Sequential Backward Selection
SD	Standard Deviation
SFR	Sequential Forward Selection
SILO	Scientific Information for Land Owners
SLFN	Single Layer Feed-forward Neural network
Step	Stepwise Regression
SVM	Support Vector Machine
SVR	Support Vector Regression
TM	Temperature model

TMBC	Bristow and Campbell Temperature Model
TMGO	Goodin Temperature Model
TMHS	Hargreaves and Samani Temperature Model
TMLI	Li Temperature Model
TRMM	Tropical Rainfall Measuring Mission
TSFS	Time Series Fourier Series
UNEP	United Nations Environment Program
UNV	Univariate Feature
UQ	Upper Quartile
VDM	Variational Mode Decomposition
VHGPR	Heteroscedatic Gaussian Processes
WEO	World Energy Organisation
WI	Willmott's Index of Agreement
XGBoost	Extreme Gradient Boosting Regression

Model Notations

The following table outline the Artificial Intelligence models developed in this study, their notations and the relevant descriptions for each Chapter of the doctoral thesis. This section is written to make it easy for any reader to quickly make any reference to the relevant models in respective Chapters and the model's input selection methodologies.

Chapter 3

Main Model	Benchmark Models	Input Selection
Artificial Neural Network (ANN)	Support Vector Regression (SVR) Gaussian Process Machine Learning (GPML) Genetic Programming (GP) Auto Regressive Moving Integrated Average (ARIMA) Temperature Model (TM) Time Series and Fourier series (TSFS)	**Feature Selection for Regression based on Neighbourhood Component Analysis (*fsrnca*)**

Chapter 4

Main Model	Benchmark Models	Input Selection
Self-Adaptive Differential Evolutionary Extreme Learning Machine (SaDE-ELM)	Extreme Learning Machine (ELM) Online Sequential ELM with Fixed Input Size (OS-ELM) Online Sequential ELM with Varying Input Size (OSVARY-ELM) PSO optimized ANN (PSO-ANN) GA optimized ANN (GA-ANN) PSO optimized SVR (PSO-SVR) GA Optimized SVR (GA-SVR)	**Ant Colony Optimization (ACO)**

[ELM optimized with Self-Adaptive Differential Evolutionary Algorithm]	Grid Search Optimized SVR (GS-SVR) Genetic Programming (GP)	

Chapter 5

Main Model	Benchmark Models	Input Selection
Three-phase PSO-W-SVR hybrid model. **[SVR optimized with PSO input selection algorithm and wavelet (moDWT) transformation]**	Artificial Neural Network (ANN) Extreme Learning Machine (ELM) Heteroscedatic Gaussian Processes (VHGPR) least-square SVR (LS-SVR) Adaptive Neuro-Fuzzy Inference System (ANFIS) Random Forest Regression (RFR) Group Method of Data Handling (GMDH) Minimax Probability Machine Regression (MPMR) Gaussian Process Machine Learning (GPML)	**Particle Swarm Optimization (PSO)**

Chapter 6

Main Model	Benchmark Models	Input Selection
Deep Belief Network (DBN)	Artificial Neural Network (ANN)	**Particle Swarm Optimization (PSO)**

Deep Neural Network (DNN)	Random Forest (RF)	**Genetic Algorithm (GA)**
	Extreme Gradient Boosting Regression (XGBoost)	Simulated Annealing (SA)
	Gradient Boosting Machine (GBM)	Stepwise Regression (Step)
	Decision Tree (DT)	Nearest Component Analysis Regression (*fsrnca*)
		Relieff Algorithm (Relieff)
		Ant Colony Optimization (ACO)
		Nondominated Sorting Genetic Algorithm (NSGA)
		Random Forest Regressor (RFR)
		Univariate Feature (UNV)
		Exhaustive Search (EXH)
		Mutual Information Regression (MIR)
		Sequential Backward Selection (SBR)
		Sequential Forward Selection (SFR)

		Recursive Feature Elimination (RFER)

Chapter 7

Main Model	Benchmark Models	Input Selection
Hybrid two-phase CLSTM [(CNN+LSTM) that integrates Convolutional Neural Networks (CNN) with the Long Short-Term Memory Networks (LSTM)]	Recurrent Neural Network (RNN) Gated Recurrent Unit (GRU) Deep Neural Network (DNN) Multi-Layer Perceptron (MLP) Decision Tree (DT)	**NONE [Antecedent lagged matrix of *GSR* time series was used as input].**

Chapter 1 Introduction

1.1 Background

Solar energy resources have become a pertinent source of cost-free electricity worldwide during the past two decades. In the last fifteen years, photovoltaic (PV) energy reached a compound annual growth rate of 40%, as reviewed in recent studies (Ghimire *et al.*, 2018). As per the World Energy Outlook (WEO) 2018, with increasing world population in coming decades, solar PV's installed capacity will surpass wind before 2025, hydropower around 2030, and coal before 2040 (Conti *et al.*, 2018). Similarly, the International Energy Agency (IEA) (Birol, 2017) has developed an optimistic scenario, according to which, electricity generation from renewable energy is expected to rise to 39% by 2050. Recently, the role of renewable energy in achieving sustainable economic growth has been the topic of interest in the global energy sector. Empowered by significant renewable energy potential, Australia's solar energy use was among the top ten nations in 2018. Australian photovoltaic power capacity reached 7.982 GW spread across 2 million installations by December 2018 (IEA, 2019) , the equivalent of more than one solar panel per person.

Furthermore, due to emerging improved technologies, solar energy extraction costs are reasonably low and large industrial-scale solar power plants provide low-cost electricity compared to fossil fuel and nuclear systems. Australia is expected to reach more than 20 GW of photovoltaic power generation in the next 20 years, equal to one-third of the current total renewable energy generation. This would support the Australian Renewable Energy Target for large-scale renewable electricity generation to reach 33,000 GWh by 2020, and for 23.5% of all electricity to come from renewable energy sources (RES) (REN21, 2018). This clearly reveals that the global market share of solar energy is expected to continue rising, and therefore, new and cost effective technologies that assess long-term energy sustainability to promote clean energy production in all parts of the World (*e.g.*, remote regions) are highly desirable.

Due to the decreasing trends of feed-in tariffs for solar PV power in many countries (including Australia), there has been an accelerated interest and need for versatile energy management schemes (EMS) for end-users to increase the generation of electricity and the capacity for power transmission from various regions, both remote and metropolitan, to meet rising consumer energy demands. EMS are able to monitor, control, and optimize the transmission and use of solar and conventional energies. However, the prediction error on the power output from a PV system can cause a negative effect on the economical profit of the system. Considering this, an accurate predictive tool for solar radiation and thus, the potential for PV power generation in a region, can help reduce the uncertainty of power generation into the future. Such tools can be used to explore and evaluate the sustainability of long-term solar powered energy installations in all regions, irrespective of their location.

Furthermore, the Australian Department of the Environment and Energy (DEE) has direct responsibility for promoting 5 out of the 17 United Nations Sustainable Development Goals (SDGs) (UN, 2015). These SDGs clearly advocate new technologies, including advanced modelling approaches to promote: (i) Goal 7 (*G7*): Affordable and Clean Energy, (ii) Goal 12 (*G12*): Responsible Production and Consumption, (iii) Goal 13 (*G13*): Climate Action, (iv) Goal 14 (*G14*): Life below Water, and (v) Goal 15 (*G15*): Life on Land. Among these five unique, yet very important goals, Goal *G7* has a direct relationship in promoting renewables and developing new and efficient energy extraction technologies. The DEE encourages and is endlessly involved in research on renewable energy related technologies (RET) to promote the SGDs. However, the stochastic and intermittent nature of any renewable resource poses numerous problems in energy security or stability needs, and this issue can limit the development of RET. These problems, associated with the stochastic nature of RES, can be eliminated, or at least, partially mitigated, by developing more precise energy prediction methodologies which are crucial for energy policy decision-makers.

The magnitude of power generated by a solar PV system is largely a function of the *GSR* and is volatile to weather conditions such as cloud motions, cloud temperature, humidity, and sunshine hours (Salcedo-Sanz *et al.*, 2018). Therefore, the knowledge and clear understanding of solar radiation availability in one particular location is a very important parameter for effective utilization of solar energy resources in design and modeling of solar energy systems, water resources management and agriculture (Wang *et al.*, 2017). Furthermore, the measured solar radiation data is not available for many of the potential sites (Khatib *et al.*, 2012; Quej *et al.*, 2017), due to the cost of the instruments (Ramedani *et al.*, 2013), improper sensor calibration and equipment failure (Díaz-Gómez *et al.*, 2015). To surmount these issues, the opportunity to adopt Reanalysis, weather station and Satellite-derived predictors to estimate short-term as well as long-term *GSR* presents an alternative and viable avenue for future exploration of solar energy.

GSR forecasting are focused on three algorithms; the first technique is based on statistical input and may include Autoregressive Moving Averaging (ARMA) (Bouzgou *et al.*, 2017), Empirical Models and Support Vector Regression (SVR) (Sarikprueck *et al.*, 2017). The second technique involves Artificial Intelligence (AI; *i.e.* Machine Learning (ML) and the much improved and recent technique based on a Deep Learning (DL[1]) algorithm), which learns from past data to build a black-box model that describes the relations between the input (predictors) and output (target) (Dayal *et al.*, 2017). These models may include the Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Neuro-Fuzzy, Genetic Programming (GP), Artificial Neural Networks (ANN) and Extreme Learning Machine (ELM). Predicted output that use ML algorithms is a popular approach summarized by many authors (*e.g.*, (Voyant *et al.*, 2017)). The third utilizes a hybrid-based approach, which employs an integration of statistical and biologically enthused methods like PSO-ANN, GA-SVR, GA-ANN, and PSO-SVR to attain accurate forecasts (Soufi *et al.*, 2017). Despite the popularity of ML and DL models for *GSR* prediction (Fentis *et al.*, 2017), in Australia the use of ML and DL has been limited. However, research into this area has been gaining more recent attention. Moreover, it has also been reported in the literature that Australia has been internationally criticised for producing very little of its energy from solar power, despite its vast resources, extensive sunshine and overall high potential (Byrnes *et al.*, 2013; Mercer, 2014).

Artificial Neural Network (ANN) has been widely used in *GSR* prediction (Marzo *et al.*, 2017) involving different meteorological and geographical parameters as the predictors. Deo and Şahin (Deo *et al.*, 2017) employed a single satellite input parameter obtained from the Earth Orbiting System - Moderate Resolution Imaging

---

[1] For the purpose of clarity, in this doctoral research thesis DL (Deep Learning) refers to an alternative form of the ML (machine learning) algorithm where a pre-defined set of multiple hidden layers including sequential, pooling and convolutional layers are incorporated for a more effective feature extraction relative to a traditional approach with a single hidden layer neuronal system (*e.g.,* an ANN model).

Spectroradiometer (EOS-MODIS), the land-surface temperature (LST) to train an ANN model and compare the *GSR* prediction performance with Multiple Linear Regression (MLR) and ARIMA models. Results showed that an ANN model outperformed both the MLR and ARIMA model. In a similar way, a Discrete Wavelet Transform (DWT) was used to decompose input metrological time-series data and combined with SVM (W-SVM) to predict *GSR* for three metropolitan cities in Australia (Deo *et al.*, 2016) . The study of Belaid et al. (Belaid *et al.*, 2016) utilized SVR. Least-square SVR using atmospheric data were utilized by Zeng *et al*. (Zeng *et al.*, 2013) to predict *GSR*. An ANN model with geographic parameters was constructed by Sözen *et al.* (Sözen *et al.*, 2004) to estimate the *GSR* in Turkey, and a further study by Alsina *et al.* (Alsina *et al.*, 2016) investigated ANN for estimation of long-term (*i.e.* monthly daily) solar radiation. Additionally, the standalone ELM (Deo *et al.*, 2019) and hybrid model in which evolutionary type algorithms integrated with ELM (Salcedo-Sanz *et al.*, 2018) were formulated to predict *GSR* using Reanalysis data for two locations in Queensland, Australia; this hybrid model outperformed the other benchmarked ML models.

Besides ANN, a plethora of studies has attempted to predict solar radiation utilising other Machine Learning approaches. Ramedani *et al.* (Ramedani *et al.*, 2014) employed a support vector regression (SVR) technique to develop a model for prediction of *GSR* in Tehran. Similarly, Gala *et al.* (Gala *et al.*, 2016) proposed a hybrid ML method by SVR, employing Gradient Boosted Regression (GBR) and Random Forest Regression (RFR) to improve the initial radiation forecasts provided by the state-of-the-art European Centre for Medium range Weather Forecasting (ECMWF) model; it has been found that the ECMWF with SVR method enhanced the solar forecasting accuracy. In addition to this, several hybrid models like GA combined with a multi-model framework (Wu *et al.*, 2014), combined Hidden Markov models and Generalized Fuzzy models (Bao *et al.*, 2013; Bhardwaj *et al.*, 2013), hybrid SVR-Wavelets (Mohammadi *et al.*, 2015), combined Self-Organizing Maps (SOM), SVR and PSO (Dong *et al.*, 2015), and a modified ANN called a Non-Linear Autoregressive Recurrent Exogenous Neural Network (NARX-NN) using recursive filtering (Hussain *et al.*, 2016) have been used to predict the *GSR*. Mohandes (Mohandes, 2012) recently developed a hybrid PSO-ANN to model longer-term monthly mean daily *GSR* values in Saudi Arabia using input parameters such as month number, sunshine duration, latitude, longitude, and altitude. The developed hybrid PSO–ANN model showed better performance compared to a Back-Propagation trained Neural Network (BP-NN).

Furthermore, Deep Neural Network, Deep Belief Network, Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) network which belong to the family of Deep Learning have shown excellent performance in a variety of applications such as computer vision, text analysis, and many others (Schmidhuber, 2015). These Deep Learning algorithms have shown remarkable results in numerous time series learning tasks such as artificial handwriting generation, language forecasting, speech recognition (LeCun *et al.*, 2015; Sutskever *et al.*, 2014), and for wind energy prediction (Chen *et al.*, 2018; Cheng *et al.*, 2017; Gers *et al.*, 2000; Liu *et al.*, 2018; Xiaoyun *et al.*, 2016). Deep Learning models, *e.g.,* LSTM, have shown a superior ability to learn long-term dependencies by maintaining a memory cell to determine which unimportant features should be forgotten and which important features should be remembered during the learning process. Therefore, by using the LSTM for modeling the *GSR*, not only can the dependence between consecutive days be captured, but the long-term (*e.g.,* seasonal) behaviour can also be learned.

For *GSR* forecasting using the Artificial Intelligence based predictive models, historical data that can be related with the target variables (*GSR*) plays a key role, also the data may not be available in all spatial regions and more importantly, in remote sites where meteorological stations are absent. Fortuitously, Reanalysis and remotely sensed data has been identified as a practical predictor for solar forecasting problems (Şenkal, 2015). In this view, the coupling of AI based models with Reanalysis and Satellite products is an improvement over station-based data as the acquisition of satellite imagery is feasible as long as a footprint is identified. Predictor variables obtained from satellite datasets are beneficial for the study of remote regions where meteorological stations are not built or are inaccessible. Satellite data is in abundance over large spatial and temporal resolutions (Qin *et al.*, 2011).

1.2 Statement of the Problem

Australia receives an average of 58 million Petajoule (PJ) of solar radiation per year, approximately 9,400 times larger than its total energy consumption of 6146 PJ (*Australian Energy Update*, 2018) in 2016-2017. Theoretically, then, if only 0.1 per cent of the incoming radiation could be converted into usable energy at an efficiency of 10 per cent, all of Australia's energy needs could be supplied by solar energy (Kent *et al.*, 2006). Moreover, the highest footprints of incident solar fluxes in Australia exist in desert regions, particularly in the northwest and centre of the continent. Despite a significant push for solar energy utilisation in outlying regions, isolation from Australia's National Electricity Market (NEM) grid is a major challenge. Additionally, in Australia, electricity networks are State controlled due to which the power plants are centrally located, therefore there are massive transmission and distribution expenses and losses (Zahedi, 2016). Natural disasters can sometimes cause considerable damage to the electricity transmission system. Recent bushfires in Western Australia destroyed up to 50 km of power lines in the South West (Tayal *et al.*, 2017). This raises the question whether alternative energies such as solar powered systems can be explored and sustainably harnessed for such isolated regions, to provide equitable access to free and affordable energy for all human populations irrespective of their geographic locations.

Considering these facts, there is great scope for harnessing solar energy, with solar energy generation being the foremost choice of energy in Australia. Almost half of the population have nominated solar power as its preferred energy source (The Climate Institute, 2017). In this context, the aim of this thesis-related research is to use the Artificial Intelligence based predictive models to provide a better performance for *GSR* prediction, because the production of the energy depends on the availability of the solar radiation, which is variable, intermittent and is not a deterministic variable due to variability of meteorological conditions. Although an accurate model is essential for *GSR* prediction, very few studies have been completed in Australia. A review of the previous studies also showed that Artificial Intelligence based predictive models ANN, SVR, ELM, GP, Gaussian Process of ML, DNN, DBN, CNN and LSTM are powerful prediction tools. However, in Australia these methods have not been explored for solar radiation prediction with Reanalysis and Satellite data as input. Additionally, with a running renewable energy target scheme in Australia, solar and wind power generating systems are being installed rapidly (Byrnes *et al.*, 2013). Therefore, the improvement of a particular Artificial Intelligence based predictive model that can predict the *GSR* with a high level of accuracy will virtually assist solar engineers, architects, agriculturists, hydrologists, and government agencies to boost the adaptation of photovoltaic electricity as a more prominent and predominant energy source into the domestic grid system.

Robust predictive models with better accuracies could serve as suitable alternatives for *GSR* prediction. In order to achieve a high prediction accuracy an optimal selection of input variables is vital. There are some features in the datasets, many of them might not be relevant to the leaning task. Some might even be noisy and their presence can increase the computational complexity and hinder the generalization capability of a prescribed model (Qi *et al.*, 2017).

Besides the input selection, in order to estimate *GSR* in a region with limited predictor dataset (inputs), a solar engineer may also be interested in checking the importance of a given predictor that effectively contributes to the *GSR* prediction process. This information can be useful for decision making in power plant design, especially in regard to selecting the most appropriate predictors and enhancing their understanding of the correct set of measurements to obtain when those data inputs are used to predict the surface level *GSR*. This study employs global sensitivity analysis using Gaussian Emulation Machine (GEM-SA) for Sensitivity Analysis software (Kennedy, 2005; Kennedy *et al.*, 2017; O'Hagan, 2006).

Other than implementing an input selection algorithm and sensitivity analysis, *GSR* (target) and the interrelated meteorological inputs exhibit seasonal characteristics, including long- and short-term fluctuations that are characterized by patterns, drifts and localized or unexpected changes in the variable. Although Artificial Intelligence based predictive models are somewhat capable of exploring non-linearities present in a model input, the accuracy of such a model is likely to be lower because of abrupt perturbations due to behaviours. To address the underlying challenges due to the presence of non-stationarities in a model's input variables, Discrete Wavelet Transform (DWT), a multi-resolution technique with a capability to decompose convoluted time-series signals into approximation (*i.e.*, high frequencies) and detailed components (*i.e.*, low frequencies), has been advocated (Zhu *et al.*, 2017) (Deo *et al.*, 2016). However, DWT can have two major disadvantages, i) the issue of decimation effect whereby half the wavelet coefficients are only used in subsequent transformation causing loss of information (Zhu *et al.*, 2014) and their dependence on the point of the commencement of wavelet transformation on input data (Rathinasamy *et al.*, 2013). Instead, a more refined and non-decimated version, known as the

Maximum Overlap Discrete Wavelet Transformation (moDWT) algorithm, can overcome these challenges (Renaud *et al.*, 2002). Considering the advantages of moDWT over conventional decomposition methods (*e.g.*, DWT), moDWT integrated with an SVR algorithm for prediction of *GSR*, has been trialed in this study.

Furthermore, considering the gaps in knowledge advocating a need for versatile tools applied in energy security devices, and also to assist in integrating solar energy variability behaviour into a real-time system, the novelty of this thesis in research is to design a new Deep Learning predictive framework based on two-layer integration of CNN and LSTM for short-term *GSR* predictions, and also to emulate the model at multi-step forecast horizons. The CNN algorithm is incorporated to extract intrinsic features of the *GSR* series, while in the second phase, LSTM is connected to CNN to utilize all relevant features for the purpose of prediction.

Overall, this doctoral thesis addresses issues of appropriate model input selection, sensitivity analysis of the model's inputs, non-linearity and non-stationarity behaviours of the model's input data in predicting the *GSR* within Australia's Solar cities. In addition, a novel two-phase hybrid CLSTM model is also explored for half-hourly *GSR* prediction to provide a near real-time simulation platform for solar energy.

1.3 Objectives

The key aim of this doctoral research thesis was to develop a set of high-precision Artificial Intelligence based predictive *GSR* models for long-term and short-term *GSR* predictions using freely available Reanalysis and Satellite data.

To achieve the key aims, the five objectives of the doctoral research thesis are outlined as follows.

- To develop ML models (ANN, SVR, GPML, GP) using Nearest Component Analysis (*fsrnca*) optimizer algorithms for predicting *GSR* at daily prediction horizon. The preciseness of the ML models was validated in respect to deterministic and statistical models. ***The article has been published in Journal of Cleaner Production (Vol. 216, Pages 288-310).***
- To utilize the hybridized ML model based on an ELM (Self Adaptive Differential Evolutionary; SaDE-ELM) algorithm to predict monthly averaged

daily *GSR* using Reanalysis and Satellite data. The input selection was done using Ant Colony Optimization (ACO), the performance of the hybrid ELM model was compared against standalone ELM, hybrid ANN and hybrid SVR model. ***The article has been published in Remote Sensing of Environment (Vol. 212, Pages 176-198).***

- To develop three-phase, ML model that utilises the SVR algorithm using the non-decimated wavelet transform (moDWT; W) and PSO optimizer algorithms for predicting monthly averaged daily *GSR* using Satellite data. Furthermore, a Gaussian Emulation Machine of Sensitivity Analysis (GEM-SA) was incorporated on screened Satellite variables to ascertain their relative role in predicting *GSR*. The preciseness of the three-phase models were validated in respect to their standalone counterparts and other popular ML models (standalone SVR, ANN, ELM, Gaussian Processes (GPML), Heteroscedatic Gaussian Processes (VHGPR), Adaptive Neuro-Fuzzy Inference System (ANFIS), Random Forest (RF), Kernel Ridge Regression (KRR), Group Method of Data Handling (GMDH) and Minimax Probability Machine Regression (MPMR)). ***The manuscript is in press for Renewable & Sustainable Energy Reviews.***
- To devise a Deep Belief Network (DBN) and Deep Neural Network (DNN) model, as a DL model with significant input interpretation capability in comparison to the conventional models (ANN, RF, Extreme Gradient Boosting Regression (XGBoost), Gradient Boosting Machine (GBM) and Decision Tree (DT)) to predict long-term *GSR* (monthly averaged daily). Fifteen diverse input selection approaches including a GEM-SA are used to select optimal MODIS-predictor variables. ***The article has been published in energies,*** **https://doi.org/10.3390/en12122407.**
- To design a new DL predictive model based on an integration of the Convolutional Neural Network (CNN) and Long Short-Term Memory Network (LSTM) algorithms tailored for short-term *GSR* predictions (half-hourly). This study also aims to emulate the CNN+LSTM (CLSTM) model at multi-step prediction horizons (1-Day, 1-Week, 2-Week and 1-Month). ***The manuscript is under review for Applied Energy.***

1.4 Thesis Layout

The thesis, presented as a collection of research publications in Scopus Quartile 1 journals, is organized into eight Chapters as follows:

Chapter 1 This Chapter presents the introductory background and the statement of problem pertaining to the research and presents the objectives of this study.

Chapter 2 This Chapter describes the study area, datasets and general methodology used in this doctoral study and also sets the scene for the following Chapters. This Chapter provides general viewpoints on the research while the specific study area, data and methods are presented in the respective following Chapters.

Chapter 3 This Chapter is presented as a published journal article in the journal, *Journal of Cleaner Production*. It is devoted to the establishment of ANN model for *GSR* prediction where the input selection uses the *fsrnca* approach. It outlines the issues with traditional approaches, model development and outcomes with respect to comparative ML model (SVR, GPML and GP) as well as deterministic models. Chapter 3 addresses the first research objective of this study.

Chapter 4 This Chapter is presented as a published manuscript in the journal, *Remote Sensing of Environment*. This Chapter describes the application of advanced optimization algorithm (SaDE-ELM) for *GSR* prediction using the Satellite and Reanalysis data. Chapter 4 is in response to the second research objective of this study. It outlines the model development and the outcomes benchmarked against standalone ELM and comparative PSO-ANN, PSO-SVR, GA-SVR and GP models.

Chapter 5 This Chapter is presented as a published manuscript in the journal, *Renewable & Sustainable Energy Reviews*. It describes the application of non-decimated wavelet transform (moDWT; W) based SVR modeling approach for *GSR* prediction using Satellite data, where the Satellite predictors are selected through PSO algorithm. Chapter 5 captures the third research objective of this study. It outlines the model development and performances of the PSO-W-SVR with respect to standalone ML, Ensemble (Gradient Boosting Machine (GBM), Extreme Gradient Boosting Regression (XGBoost) and Decision Tree (DT)) and statistical models.

Chapter 6 This Chapter is presented as a published manuscript in *Energies*. It presents the development of a Deep Learning prediction technique for *GSR* prediction. It outlines the model development and performances of the DBN and DNN model with respect to Single Hidden Layer (*i.e.*, Artificial Neural Network) and Ensemble models (Random Forest Regression, GBM, XGBoost and DT). Chapter 6 addresses the fourth research objective of this study.

Chapter 7 This Chapter is presented as a published manuscript in *Applied Energy* (Under second stage review). It presents the development of a two-phase hybrid CLSTM model for *GSR* prediction (half-hourly). It outlines the model development and performances of the CLSTM model with respect to Two-phase CLSTM hybrid model, benchmarked with competing approaches (*i.e.*, CNN, LSTM, DNN, Recurrent Neural Network (RNN), Gated Recurrent unit Neural Network (GRU)), Single Hidden Layer Neural Network using Multilayer Perceptron's (MLP) and DT). Chapter 7 addresses the fifth research objective of this doctoral study.

Chapter 8 This Chapter presents the synthesis of the study with concluding remarks, limitations, and recommendations for future works.

Chapter 2 Data and Methodology

This Chapter provides an overview of the location of the study sites used in developing the Artificial Intelligence (AI) and Deep Learning (DL) based *GSR* predictive model. Note that the DL models refer are a special category of the AI models with a more sophisticated learning algorithm developed in this study. Different study sites within Australia were selected to achieve each objective, which is described in detail in each of the Chapters. The description of data used, length of data and limitations if any, are also presented. Although the methodology and model development are well described in each Chapter, the brief account of methodology is also discussed in this Chapter. The description of the study area is given next. This is followed by the data used and the general procedure used in this doctoral research thesis for development of Artificial Intelligence based *GSR* predictive models.

2.1 Study Area

In order to design the predictive model for the *GSR* prediction, Australia's solar cities(Beatley, 2007; Kuwahata *et al.*, 2011): Alice Springs (Northern Territory), Coburg (Victoria), Perth (Western Australia), Adelaide (South Australia), Blacktown (New South Wales), Central Victoria (Victoria) and Townsville (Queensland) were selected. In addition to these solar cities, precise *GSR* model for few highly populated cities like Brisbane, Sunshine Coast, Ipswich, Gold Coast and Toowoomba from south east Queensland (SEQ) were also developed. These regions, which are naturally solar-rich due to their distinct geographic location, present a unique case for testing the Artificial Intelligence (AI) based predictive models. For example, Alice Springs, with a population of 28,000, has a desert climate (*BWh*) as per the Köppen climate classification (Belda *et al.*, 2014; Lohmann *et al.*, 1993) occupying about 9% of the territory population where the climate is reflective of semi-arid conditions that have relatively hot summer temperatures. A mean daily maximum temperature above 30 °C occurs for six months of every year with almost 300 days of sunshine per year and during a typical sunshine day receiving approximately 1 kW of solar energy from the Sun (Havas *et al.*, 2015; Linacre *et al.*, 1997) .On the other hand, Coburg has a temperate oceanic climate (*Cfb*) with the maximum temperature ranging from 32 °C in summer to 15 °C in winter. All these study sites show disparate geophysical features in terms of primary climate classes and differing elevations. Further, SEQ region have sub-tropical climate influenced by tropical systems from the north and fluctuations in the high-pressure ridge to the south. SEQ is one of Australia's largest and fastest growing urban regions, with the population concentrated along the coast (Dedekorkut *et al.*, 2010; Helfer *et al.*, 2012; Mantyka-Pringle *et al.*, 2014). Recent data from Clean Energy Council (Council, 2018) shows that until January 2019 the total installed capacity of roof top solar photovoltaics (PV) in SEQ was 1451 MW (Council, 2018), more than 450,000 household have solar PV installed (SunWiz, 2019). In this region an accurate model is required for the *GSR* prediction so that the installed technology can be operated at maximum efficiency. In a concise way, Figure 2.1 shows the map of each location for development of Artificial Intelligence based *GSR* predictive models in achieving each objective.

2.2 Data Description

A variety of data sources were utilized in developing high precision Artificial Intelligence based predictive *GSR* models. To summarize, Table 2.1 provides the details of the data used with respective sources and other relevant details in achieving each objective.

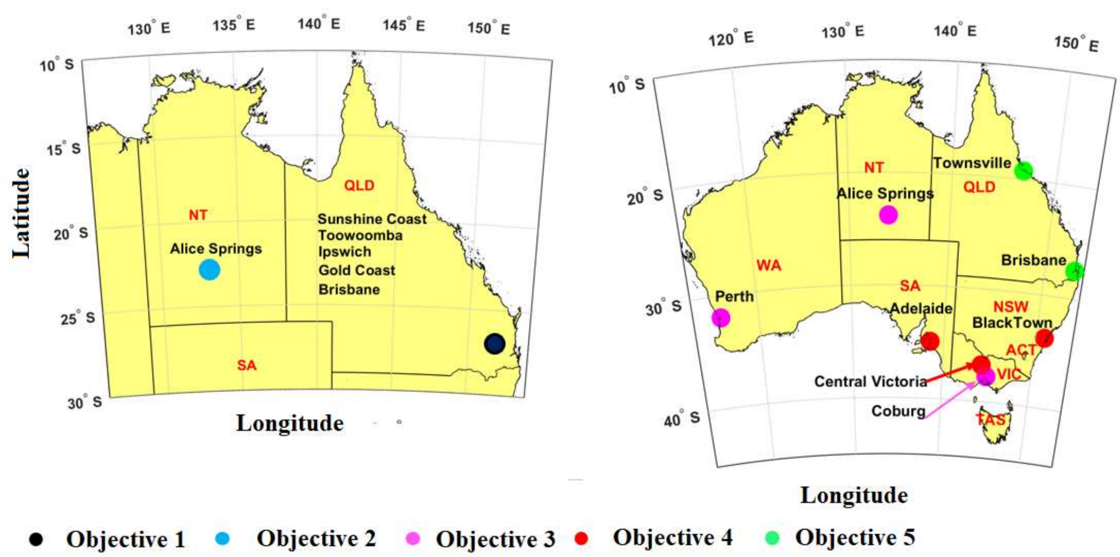


Figure 2.1 Geographic location of Australia's solar cites and south east Queensland, illustrating the selected study sites for Objectives 1 to 5, respectively.

Table 2.1 Details of all data used in this study.

Objective	**Location**	**Data Source**	**Data Period**	**Prediction Horizon**
1 (Chapter 3)	Brisbane, Ipswich, Gold Coast, Toowoomba, and Sunshine Coast	Predictors: Atmospheric parameters from European Centre for Medium range Weather Forecasting (ECMWF) (Dee *et al.*, 2011) and Meteorological variables from Scientific Information for Land Owners (SILO). Target: Daily *GSR* data from SILO	01 January 1979 to 01 December 2016	Short Term (daily)
2 (Chapter 4)	Brisbane and Townsville	Predictors: Atmospheric parameters from ECMWF and Remotely sensed atmospheric parameters from Moderate Resolution Imaging Spectroradiometer (MODIS)(Zhou *et al.*, 2019). Target: Daily *GSR* data from SILO	01 March 2001 to 01 August 2017	Long term (monthly mean daily)

3 (Chapter 5)	Alice Springs, Coburg, and Perth	Predictors: Remotely sensed atmospheric parameters from MODIS. Target: Daily *GSR* data from SILO	01 March 2000 to 01 August 2017	Long term (monthly mean daily)
4 (Chapter 6)	Central Victoria, Black town, Adelaide, and Townsville	Predictors: Remotely sensed atmospheric parameters from MODIS. Target: Daily *GSR* data from SILO	01 March 2000 to 01 August 2018	Long term (monthly mean daily)
5 (Chapter 7)	Alice Springs	Predictors: Antecedent half hourly *GSR* data from Bureau of Meteorology (BOM). Target: Half-hourly *GSR* data from BOM.	01 January 2006 to 27 August 2018	Short Term (half-hourly)

2.2.1 Meteorological data - Scientific Information for Land Owners and Bureau of Meteorology (BOM).

All target variable data (*GSR*) for Chapter 1, Chapter 2, Chapter 3 and Chapter 4 were acquired from Scientific Information for Land Owners (SILO) database: https://www.longpaddock.qld.gov.au/silo/ppd/index.php. SILO is a database system designed to provide users of biological and hydrological models 'ready-to-use' climate data. In SILO-database missing values had been interpolated with robust statistical tools applied in the quality control stages implemented by the Australian Bureau of Meteorology (BOM) (Beesley *et al.*, 2009; Zajaczkowski, 2009). This has empowered SILO-based Meteorological data to be employed in previous solar energy-related studies (Deo *et al.*, 2016; Salcedo-Sanz *et al.*, 2018). Furthermore, in this doctoral research thesis, following recommendation of Simmons *et al.* (Simmons *et al.*, 2010) the meteorological data (Maximum Temperature, Minimum Temperature, Relative Humidity, Rainfall, Evaporation and Vapour Pressure) from SILO-database are integrated with Reanalysis data as predicants for developing Artificial Intelligence based *GSR* predictive models (Chapter 3 and Chapter 4). A complete list of input parameters used is provided in Chapter 3 (Table 2) and Chapter 4 (Table 2). These SILO-based meteorological data were produced by the Queensland Climate Change Centre of Excellence (QCCCE) within the Queensland Department of Environment and Resource Management and is available commercially as patched point records and as a synthetic dataset generated over a set of evenly spaced grid locations, referred to as the 'drilled data'.

To achieve objective 5, one minute *GSR* data for Alice Springs were acquired from the Australian Bureau of Meteorology **(**http://reg.bom.gov.au/ climate/reg/oneminsolar/ index.shtml**)**. *GSR* is estimated from three sources (Zajaczkowski *et al.*, 2013):

- Radiometer data: direct solar radiation measurements, quality controlled data available 1999-current.
- Sunshine duration measurements in hours from sunshine recorders.
- Cloud cover observations in oktas: okt9 and okt15 (in 0–8 scale) corresponding to observations at 0900 and 1500. A smaller subset of stations records cloud-cover also at other times of day.

This doctoral research thesis utilizes time series of *GSR* (over preceding 1 minute intervals) from 01-January-2006 to 27-August-2018 for Australian Bureau of Meteorology, Station ID: 015590 (Alice Springs Airport; Lat.-23.79 ^{o}S, Long. 133.89 ^{o}E). Notably, *GSR* measurements have been performed simultaneously, 24 hours a day, at equidistant time intervals of 1 minute. Only the data from 07:00 AM to 06:00 PM over a 30-minute interval are used for designing CLSTM hybrid predictive model as these times represent a period of meaningful daylight hours. The Australian Bureau of Meteorology uses matched sensors for diffuse and global pyranometry, and instruments are chosen such that there is a 95% confidence that there will be < 1% change in sensitivity over 12 months due to sensor degradation. However, pyranometer sensitivity may change with time and exposure to radiation, mainly due to the deterioration of the sensor (BOM, 2019).

2.2.2 Atmospheric parameters - Interim ERA European Centre for Medium Range Weather Forecasting (ECMWF) Reanalysis

To achieve objective 2 and 3, a total of 87 possible predictor inputs were acquired from ERA- Interim (Reanalysis). A complete list of input parameters used is provided in Chapter 3 (Table 2) and Chapter 4 (Table 2). The ERA-Interim Reanalysis data is released by European Centre for Medium-Range Weather Forecasts (ECMWF) (Dee *et al.*, 2011) . ERA-Interim is the most comprehensive set of assimilation satellite observation data in Reanalysis data (You *et al.*, 2019). ERA-Interim is generated using the ECMWF Integrated Forecasting System (IFS) Cy31r2 model and four-dimensional variational data assimilation (4D-Var). ERA-Interim replaces the 1990's and 2000's ERA-40 Reanalysis data and provides improved atmospheric model and assimilation system with high resolution 1.5° × 1.5° latitude–longitude grids with 37 pressure levels (http://data.ecmwf.int/data) (Akhil Raj *et al.*, 2015). It includes an "interim" Reanalysis of the period 1979 to the present time and updates in near real time (Cheng *et al.*, 2014); four times per day at 00, 06, 12, and 18 UTC (Reis *et al.*, 2015). In addition to the effects of instrument and calibration errors, biases in in radiance data assimilation are affected by systematic errors in the radiative transfer models (RTTOV-7) that are embedded in the assimilation system. In order to cope with is bias error, in ERA-Interim, the estimation of bias parameters for satellite radiance data is handled automatically by a variational bias correction system. An important practical advantage of this approach is that it removes the need for manual tuning procedures, which are prone to error (Dee *et al.*, 2009).

Moreover, the introduction of new "Wavelet" like weighting functions (J_b) (Fisher, 2003) to cope with background error covariance and the utilization of rain-affected radiances rather than derived rain rates for rainfall assimilation are further enhancements in Reanalysis data. Nevertheless, uncertainties and biases in these Reanalysis data are very difficult to quantify; it is therefore recommended to consider reanalysis data in tandem with the more traditional, observation-only climate datasets (Simmons *et al.*, 2010). Thus, in this doctoral research thesis Reanalysis data are integrated with ground-based weather station data from SILO to develop the Artificial Intelligence based *GSR* predictive models (Chapter 3 and Chapter 4).

2.2.3 Atmospheric Parameters - The Moderate Resolution Imaging Spectroradiometer (MODIS)

The use of satellite-based data overcomes the limitations of site measurements and provides an alternative for obtaining the spatial distribution of solar radiation (Ibrahim *et al.*, 2017; Quesada-Ruiz *et al.*, 2015). Hence, for this reason in this doctoral research thesis for objective 2, 3 and 4 the satellite based atmospheric parameter (MODIS) are used to develop Artificial Intelligence based predictive *GSR* model. MODIS is an instrument aboard the Terra (EOS AM) and Aqua (EOS PM) satellites. Terra's orbit around the Earth is timed so that it passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon. Terra and Aqua cover the globe every 1-2 days, providing data in moderate spatial resolution (250 m at nadir) with wide swath (2330 km) and large spectral range (36 channels between 0.412 and 14.2 μm)(López *et al.*, 2014). Forty-four data products are retrieved from the MODIS observations. Among these products, the MOD08-M3 contains approximately 800 sub-datasets describing features of the atmosphere, such as the cloud fraction, cloud optical thickness, precipitable water vapor amount, and aerosol optical thickness (Kim *et al.*, 2010). These remotely sensed atmospheric products can be considered as an alternative predictor for Artificial Intelligence based *GSR* predictive model, particularly for the remote locations with no ground-based measurement infrastructure.

Additionally, for long-term forecast horizons (*e.g.*,, monthly), satellite data remains in abundance for a diverse range of spatial and temporal resolutions, and recently, have been adopted in global solar radiation prediction problems (Deo *et al.*, 2017; Deo *et al.*, 2019). Although recent studies have considered solar radiation models trained with MODIS datasets, these were limited to cloud free predictor variables and land surface temperature (LST). Considering this, MODIS-M3 product has not been conducted and models have not been developed and investigated in Australia. Therefore, exploring new models using the MODIS-M3 product is of vital importance and significance.

To design an Artificial Intelligence based *GSR* predictive model over long-term horizons, monthly predictor data (MODIS-M3) have been extracted from Goddard Online Interactive Visualization and Analysis Infrastructure (Giovanni) repository (https://giovanni.gsfc.nasa.gov/giovanni). This Giovanni repository is maintained by National Aeronautics and Space Administration (NASA).

Table 1 (Chapter 4), Table 2 (Chapter 5) and Table 1 (Chapter 6) list the predictors acquired from Giovanni.

2.3 General Methodology

Prior to the model development, data quality checking phase is necessary. Due to equipment faults or site closure in a period, there have been some missing data, filled with mean value of previous years as a common practice. For instance, in Australian Bureau of Meteorology database, 20% of *GSR* data for March (2018) were missing and filled with mean values from March months of 2006 – 2017. Furthermore, the meteorological data and the interrelated atmospheric parameters, as well as the climatic indices, naturally display stochastic behaviour. In addition, the inputs are in the different set of units or are dimensionless. As a result, the appropriate scaling is required to avoid the dominance of inputs with large numeric ranges that in turn may undermine the effects of lower range values. Normalization also brings the data to a common scale and avoid extra iteration during model learning process. Therefore, all predictor inputs and the target were normalized to the range of zero and one using the below equation Eq. (1) and then returned to the original values after the simulation by application of Eq. (2).

$$X_n = \frac{X_{actual} - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

$$X_{actual} = X_n \left(X_{\max} - X_{\min} \right) + X_{\min} \tag{2}$$

where X, X_{min} and X_{max} represent the input data value and its overall minimum and maximum values, respectively.

In this doctoral research thesis, various Artificial Intelligence based predictive models are considered for an evaluation of their preciseness in emulating *GSR* since a robust modeling approach is necessary. The models range from the well-known neuronal Machine Learning (ML) ANN, ELM, SVR (Chapter 3 to 5) model to the more efficient and advanced Deep Learning (DL) DBN, DNN, CNN and LSTM (Chapter 6 to 7).

ANN can be identified as a simplified mathematical model based on the neurological structure of human brain. The ability of ANN in determining complex relationships among variables makes this technique one of the most powerful tools in data modeling field (Akbari *et al.*, 2014). The basic unit in an ANN is the neuron (node). Neurons are connected to each other by links known as synapses, associated with each synapse there is a weight factor (Antonopoulos *et al.*, 2019). The ELM model is also neuronal algorithm like ANN but the ELM utilizes a Single Layer Feed forward Neural Network (SLFN) to learn the pertinent predictive features from the historical data. ELM is comparatively faster and computationally convenient in relation to the ANN and SVR (Akbari *et al.*, 2014; Blanchard *et al.*, 2018; Lunsford *et al.*, 2019; Song *et al.*, 2017; Spolaore *et al.*, 2017). Similar to ANNs, SVR use an implicit feature space mapping from the dimension of the data to a possibly infinite feature space, providing a non-linear representation of the modeled data; this is done through the 'kernel trick'(Akbari *et al.*, 2014; Dhiman *et al.*, 2019). The SVR model has been accepted as a universal tool for solving multidimensional function estimation problems.

Further, Deep Learning model like DNN, DBN and CNN uses neural networks structures to represent the data. The concept of DNN is closely associated with ANN with many hidden layers and nodes in each layer (Liu *et al.*, 2017) and is able to learn a set of features that will be later used in order to approximate the objective function. Similarly, CNN is variant of DNN consisting of one or more convolution, pooling and fully connected layers (Wang *et al.*, 2019). Each convolutional layer consists of several convolutional units, and parameters of every unit are optimized by a back-propagation algorithm. The purpose of a convolutional manipulation in CNN is to extract unique features of the input layer. The DBN is a probabilistic, generative model that can learn to probabilistically reconstruct its inputs and is composed of multiple simple learning modules. The main aim of DBN is the weight initialization of a deep neural network to produce optimum model in comparison to the model by random weights (Ghasemi *et al.*, 2018).

As a distinctive class of Recurrent Neural Network, LSTMs utilize special units named memory blocks to take the place of the traditional neurons in the hidden layers (Hochreiter *et al.*, 1997; Sainath *et al.*, 2015). Moreover, there exist three gates units called input gates, output gates and forget gates in memory blocks and hence LSTMs have the ability to update and control the information flow in the block through these gates (Chen *et al.*, 2018).

In Chapter 7 of this doctoral research thesis, the CLSTM model is proposed, in this CLSTM model, the CNN algorithm is incorporated to extract intrinsic features of the *GSR* time series, while in the second phase, LSTM is connected to CNN to utilize all relevant features for the purpose of prediction.

In order to handle the non-stationarity features within the inputs, data pre-processing via proper multi-resolution analysis tool is necessary. Hence, hybridized models with advanced non-decimated wavelet Multi-Resolution Utility (moDWT, W) is adopted (Chapter 5). In addition, new approaches are developed and explored including an ACO_SaDE-ELM, three phase SVR model (PSO_W_SVR) and CLSTM (CNN_LSTM) model.

Appropriate input selection is imperative not only for input dimension reduction but also to improve the model performances. The optimization by means of input selection approaches also has its own advantages and disadvantages and therefore many algorithms were explored including the Neighborhood Component Analysis (NCA) based input selection algorithm for regression (*fsrnca*), ACO, PSO, GA and RF algorithm. Other than incorporating the input selection strategy, sensitivity analysis was also done to examine the statistical relationships between *GSR* and its selected input variables using GEM-SA.

Table 2.2 summarizes the details of the methodology and tools used for the development of Artificial Intelligence based *GSR* predictive models. The specific models developed in this study include:

1. Four ML models ANN, SVR, GPML and GP for daily *GSR* prediction. "*fsrnca*" was utilized for input optimization *(*Chapter 3*)*.
2. Hybrid ML model; SaDE-ELM hybrid model is designed by integrating evolutionary algorithm with SaDE for optimization of neuronal hidden-layer weight. ACO algorithm is then incorporated in the SaDE-ELM to screen the appropriate predictors in accordance with their functional relationship with *GSR (*Chapter 4*)*.
3. Three phase ML model PSO-W-SVR, for monthly averaged daily *GSR* prediction was developed. PSO was utilized for input optimization with moDWT (W) for addressing non-stationarity. Further GEM-SA was utilized for the sensitivity analysis of selected predictors *(*Chapter 5*)*.
4. DL models (DBN and DNN) for monthly averaged daily *GSR* prediction. A total of 5 filter- and 10 wrapper-based input selection algorithms are employed. Further GEM-SA was utilized for the sensitivity analysis of selected predictors *(*Chapter 6*)*.
5. New two-phase DL model (CLSTM) for half-hourly *GSR* prediction. The Autocorrelation Function (ACF) and Mutual Information Test were utilized for determination of significant lags *(*Chapter 7*)*.

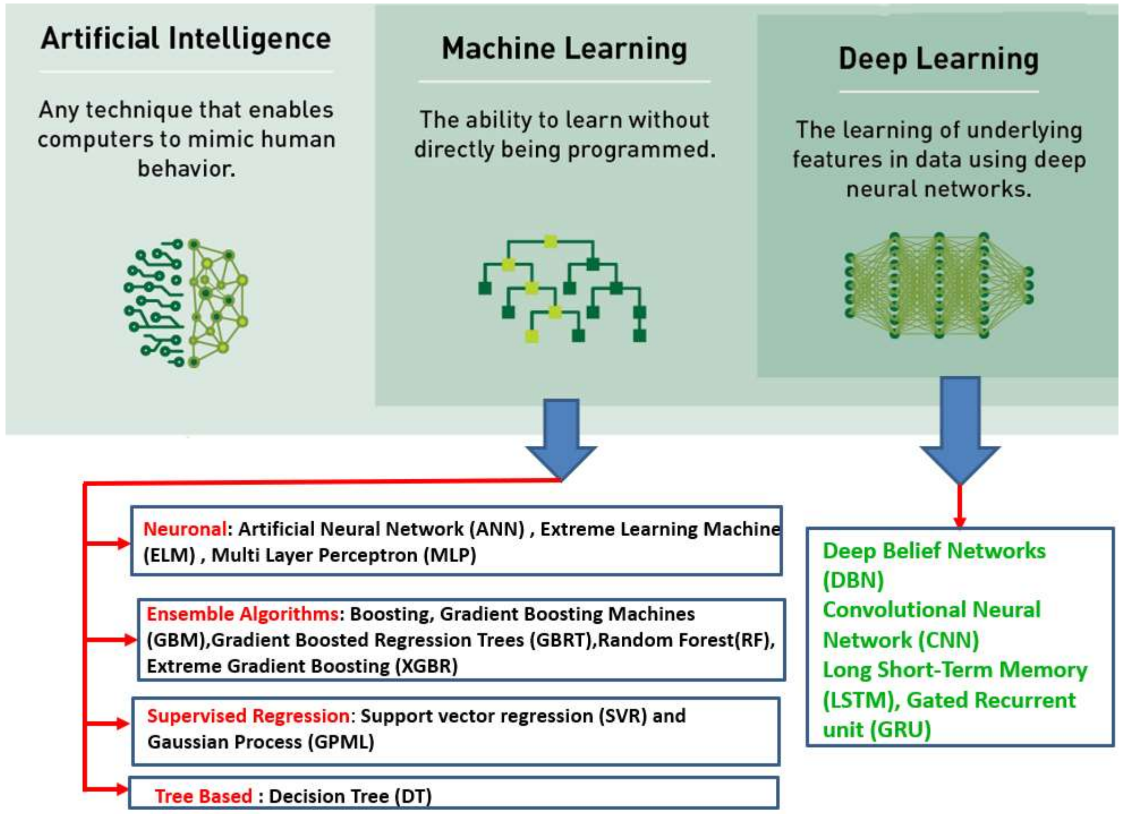


Figure 2.2 Brief overview of Artificial Intelligence (AI) based predictive *GSR* models used in this doctoral research thesis.

Table 2.2 Summary of the methodology and tools used for the development of Artificial Intelligence based *GSR* predictive models.

Objective	Main Model	Benchmark Model	Data Pre-processing	Programming Tools used
1 **(Chapter 3)**	ANN	SVR, GPML, GP, ARIMA, TM, TSFS	Normalization, cross-correlation, and Feature selection using '*fsrnca*' algorithm	MATLAB, Minitab

2 (Chapter 4)	SaDE-ELM	ELM, OS-ELM, OSVARY-ELM, PSO-ANN, GA-ANN, PSO-SVR, GA-SVR, GS-SVR	Normalization and Feature selection using ACO algorithm	MATLAB, Minitab
3 (Chapter 5)	SVR	ANN, ELM, VHGPR, LS-SVR, ANFIS, RF, GMDH, MPMR, GPML	Normalization, decomposition of input using moDWT process, Feature selection using PSO algorithm and sensitivity analysis of predictors	MATLAB, R-software
4 (Chapter 6)	DBN, DNN	ANN, RF, XGBoost, GBM, DT	Normalization and Feature selection using 16 different Input Selection (IS) algorithms and sensitivity analysis of predictors	Python, MATLAB, R-software
5 (Chapter 7)	CLSTM	RNN, GRU, DNN, MLP, DT	Normalization, Autocorrelation and Mutual Information Test.	Python, MATLAB, R-software

2.3.1 Model Evaluation

To evaluate the performance of Artificial Intelligence based predictive *GSR* models, several statistical metrics were employed. They were based on Root Mean Square Error (*RMSE*), Mean Absolute Error (*MAE*), Correlation Coefficient (*r*) Willmott' s Index (*WI*), Nash–Sutcliffe Coefficient (E_{NS}) and the Legates and McCabe Index (E1). Furthermore, relative (%) error values based on the *RMSE* and *MAE* are also used for model comparison at geographically distinct sites. Additionally, in Chapter 5, 6 and 7 the Kling-Gupta Efficiency (*KGE*) (Gupta *et al.*, 2009) and Absolute Percentage Bias (*APB*) metrics was also computed over the tested data. Besides these statistical metrics, the Artificial Intelligence based predictive *GSR* models are also analyzed with diagnostic plots including box plots, scatter diagram, histogram, time series plot, spider plot, Lowry Plot and Taylor plots.

2.3.2 Software Package and Tools

In this doctoral research thesis, the Artificial Intelligence based *GSR* predictive models (*i.e*., Machine Learning & Deep Learning) are developed under Intel core *i*7 @ 3.3 GHz and 16 GB memory computer. For the model construction, *MATLAB* (MathWorks, 1996) (Chapter 1, 2 and 3) and *Python Software* (Sanner, 1999) (Chapter 4 and 5). Freely available libraries based on Deep Learning (DL) abilities (*i.e*., *Keras* (Ketkar, 2017), *Tensor Flow* (Abadi *et al.*, 2016) & *Sklearn* (Pedregosa *et al.*, 2011)), have been used. Other programming tools such as *Minitab* (Ryan, 1994) (Chapter 1,2 and 3) is used for statistical analysis of modelling data and *R-Software* (Benoit *et al.*, 2018) is used for Lowry Plot.

Chapter 3 Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cities of Queensland Australia

Foreword

This Chapter is an exact copy of the published article in *Journal of Cleaner Production* **216** (2019) 288-310. Scopus Impact Factor 6.680.

This article demonstrates the applicability of ANN (ML) model for the daily *GSR* prediction. Four different Machine Learning models 1) Artificial Neural Network (ANN), 2) Genetic Programming (GP), 3) GPML (Gaussian Process of Machine Learning (GPML) and 4) Support Vector Regression (SVR) and 7 deterministic (Temperature Model (TM) Time Series with Fourier Series (TSFS) and ARIMA) models; totally 11 different models evaluated, providing an extensive validation of ANN for daily *GSR* prediction. The eighty five inputs from Reanalysis and SILO are screened using a two phase Input Selection (IS) method. Firstly, the cross correlation between input and target followed by the Neighbourhood Component Analysis (NCA) based input selection algorithm for regression purposes (*fsrnca*) is applied to determine the relative feature weights.

Additionally, the ANN model for daily *GSR* prediction is verified for seasonal and large-scale climatic irregularities (*e.g.*, ENSO and IOD) for SEQ study site.

Chapter 4 Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed MODIS satellite and Reanalysis atmospheric products in solar-rich cities

Foreword

This Chapter is an exact copy of the published article in *Remote Sensing of Environment* **212** (2018) 176–198. Scopus Impact Factor 8.100.

The prediction of monthly average daily *GSR* is undertaken in this Chapter, by employing the self-adaptive differential evolutionary Extreme Learning Machine (Hybrid, SaDE-ELM). The self-adaptive differential evolutionary algorithm (SaDE) was applied for optimization of neuronal hidden-layer weight of ELM model. Sixty-seven predictor variables were sourced from Giovanni and Eighty seven predictor variables from Reanalysis. The metaheuristic input selection algorithm called Ant Colony Optimization (ACO) was used to select the most important 20 predictor variables to forecast the monthly daily average *GSR* for Brisbane and Townsville.

The SaDE-ELM is then benchmarked with nine different ML models: a basic ELM, genetic programming (GP), online sequential ELM with fixed (OS-ELM) and varying (OSVARY-ELM) input sizes, and hybridized model including the Particle Swarm Optimized-Artificial Neural Network model (PSO-ANN), Genetic Algorithm optimized ANN (GA-ANN), PSO-Support Vector Machine model (PSO-SVR), Genetic Algorithm optimized-SVR model (GA-SVR) and the SVR model optimized with Grid Search (GS-SVR).

Furthermore, the prediction capability of SaDE-ELM model is tested for large scale climatic anomalies (*e. g.*, ENSO and IOD) for Brisbane and Townsville.

Chapter 5 Wavelet-based 3-phase hybrid SVR model trained with particle swarm optimization and maximum overlap discrete wavelet transform for solar radiation prediction with remote sensing satellite-derived predictors

Foreword

This Chapter is an exact copy of the published manuscript in the *Renewable and Sustainable Energy Reviews*, 113 (2019) 109247. Scopus Impact Factor 10.49.

This Chapter describes the hybridization of the widely used regression based Machine Learning model, Support Vector Regression (SVR) for monthly averaged daily *GSR* prediction for three solar cities of Australia (Alice Springs, Coburg and Perth).

To acquire relevant model input features, Satellite derived (MODIS) variables are screened with the Particle Swarm Optimization (PSO) algorithm, and a Gaussian Emulation method of sensitivity analysis (GEM-SA) is incorporated on all screened variables to ascertain their relative role in predicting *GSR*. To address pertinent issues of non-stationarities, PSO selected variables are decomposed with Maximum Overlap Discrete Wavelet Transformation (moDWT; W) prior to its incorporation in SVR, constructing a three-phase PSO-W-SVR hybrid model where the hyper-parameters are acquired by evolutionary (*i.e.*, PSO & Genetic Algorithm) and Grid Search methods.

The three phase Machine Learning model (PSO-W-SVR) is evaluated against the comparative ANN, Extreme Learning Machine (ELM), Gaussian Processes regression (GPR), Heteroscedatic Gaussian Processes (VHGPR), least-square SVR (LS-SVR), Adaptive Neuro-Fuzzy Inference System (ANFIS), Random Forest (RFR), Group Method of Data Handling (GMDH) and Minimax Probability Machine Regression (MPMR) models.

Chapter 6 Deep Learning neural network trained with MODIS satellite-derived predictors for long-term global solar radiation prediction

Foreword

This Chapter is an exact copy of the published article in *Energies* 2019,12(12) 2407. Scopus Impact Factor 2.670.

Compared with other Machine Learning models, the Deep Learning (DL) models can extract the deep inherent features in a dataset. Hence in this study, two algorithms based on Deep Belief Network (DBN) and Deep Neural Networks (DNN), as popular DL models with feature interpretation capability in comparison in respect to Machine Learning models, are explored to purposely predict long-term *GSR (monthly averaged daily)* for four solar cities of Australia (Adelaide, Blacktown, Townsville, and Central Victoria).

Five filter based and ten wrapper based Input Selection (IS) methods is used to extract the important predictor variables from Satellite data (MODIS), further sensitivity analysis is done using the Gaussian Emulation Machine of sensitivity analysis (GEM-SA).

The DBN and DNN is evaluated against the comparative Artificial Neural Network (ANN) and Ensemble models [Random Forest Regression (RF), Extreme Gradient Boosting Regression (XGBoost), Gradient Boosting Machine (GBM) and Decision Tree (DT)].

Chapter 7 Deep Solar Radiation Forecasts with Hybrid Convolutional Neural Network integrated with Long Short-term Memory Network Algorithms

Foreword

This Chapter is an exact copy of the published manuscript to the *Applied Energy* 253 (2019), 113541. Scopus Impact Factor 8.57.

In this Chapter, real time prediction is done using two-phase hybrid Deep Learning (DL) model. A new model by integrating Convolutional Neural Network (CNN) with Long Short-Term Memory Network (LSTM) is developed (CLSTM) to predict half-hourly *GSR* for Alice Springs. In this CLSTM model, CNN is used to extract *GSR* data features and LSTM to encapsulate the features to generate a low latency-based time series *GSR* prediction. Minute level *GSR* data for Alice Springs are extracted, stationarity checks applied via unit-root test. Further, in order to construct the input matrix of antecedent *GSR* values, Auto Correlation and mutual information test was applied.

The two-phase hybrid CLSTM model is benchmarked with standalone model (CNN and LSTM) along with Deep Neural Network (DNN), Recurrent Neural Network (RNN), Gated Recurrent unit Neural Network (GRU), Multilayer Perceptron (MLP) and Decision Tree (DT).

Additionally, the two-phase hybrid CLSTM hybrid predictive model is tested for a 1-Day forecast horizon over a full diurnal cycle (*i.e.*, 23 test points), 1-Week forecast horizon over a 7-day period (161 points), 2-Week forecast horizon over a 14 day period (322 points) and 1-Month forecast horizon over a 30 day period (621 points). The model is also tested over 2, 3, 4, 5, 6, 7 and 8-Monthly horizon.

Chapter 8 Synthesis, Conclusions and Future Scope

8.1 Synthesis and Conclusions

Solar power is a vast, free and renewable resource that can be used to produce electricity. Solar energy is a commercially-proven, rapidly growing form of electricity generation. Accurately predicting solar radiation can help enhance financial efficiency and acceptability of solar energy generation and utilization. In this study, the feasibility of Artificial Intelligence based predictive model for predicting long-term and short-term global solar radiation (*GSR*) was investigated. In order to develop the predictive model data were acquired from the Scientific Information for Land Owners (SILO), European Centre for Medium Range Weather Forecasting (ECMWF) Reanalysis and Moderate Resolution Imaging Spectroradiometer (MODIS; satellite data). Artificial Intelligence (AI) based *GSR* Predictive models are validated for Solar cities of Australia (Alice Springs, Coburg, Adelaide, Central Victoria, Perth, Townsville, and Blacktown) and five location of South east Queensland (SEQ) (Brisbane, Ipswich, Gold coast, Sunshine Coast and Toowoomba). The prediction interval was varied from monthly averaged daily (long-term) to daily and half-hourly (short-term). The Machine Learning (ML) and Deep Learning (DL) algorithms that were utilized to design Artificial Intelligence based predictive models included, Artificial Neural Network (ANN; ML), Extreme Learning Machine (ELM; ML), Support Vector Regression (SVR; ML), Long Short-term Memory (LSTM; ML), Deep Neural Network (DNN; DL), Deep Belief Network (DBN; DL) and the Convolutional Neural Network (CNN).

Four important issues were addressed in this doctoral study (i) the problem of selection of non-redundant predictor inputs from sets of multivariate input in *GSR* prediction, (ii) sensitivity analysis of selected predictors, (iii) non-stationarity and non-linearity issue and (iv) use of Deep Learning model in *GSR* prediction study. Several filter and wrapper based algorithms including swarm based (Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO)) resolved the first issue of feature optimization by removing the irrelevant features. Further, the sensitivity analysis of the relevant inputs are done using Gaussian Emulation Machine of Sensitivity Analysis (GEM-SA) to ascertain their relative role in predicting *GSR*. The issue with non-stationarity was resolved by the use of non-decimated Discrete Wavelet Transform (DWT). Finally, the standalone DL model (DBN, DNN) for long-term prediction and two-phase hybrid CNN+LSTM model for real time (half-hourly) prediction.

In the first objective (Chapter 3), the Neighborhood Component Analysis (NCA) based Input Selection (IS) algorithm for regression (*fsrnca*) was employed to extract the optimal features, validated against ANN, Gaussian Process of Machine Learning (GPML), Support Vector Regression (SVR), Genetic Programming (GP) and other deterministic models to emulate the estimation of daily *GSR* at the five locations of SEQ. The 20 most important input variables were chosen from 85 inputs (from Reanalysis and SILO) and by integrating these most important predictor variables from January 1979 to June 2009. This study has constructed and authenticated the ANN model and compared its overall performance with other ML models like SVM, GP and GPML. The result shows that an ANN model is able to model the causative relationship between meteorological data and solar radiation and this study can be utilised by interested stakeholders, including solar energy investors and Engineers to predict the solar radiation for sites where ground-based stations are not available. It can also be applied for site selection and prioritization purposes to assess whether the project is economically and monetarily feasible in terms of solar investment. Additionally, the sensitivity analysis confirms that there was shared effect of predictors on the target variable (*GSR*), for instance in Brisbane, the minimum 18 predictors are essential to achieve the Relative Root Mean Square Error (*RRMSE*) error below 10%. Also, when ANN models are assessed with predictors grouped into El Niño, La Niña (ENSO) and the positive, negative and neutral periods of Indian Ocean Dipole (IOD), affirmed the merits of ANN model ($RRMSE \leq 11\ \%$). Seasonal analysis showed that ANN was an elite tool over SVR, GPML and GP for *GSR* prediction.

The hybrid model was proposed in second objective (Chapter 4) for the prediction of long-term (monthly averaged daily) *GSR* using the MODIS and Reanalysis data for solar rich cities of Queensland (Brisbane and Townsville). The ELM model hidden neuron weights were optimized using the Self Adaptive

Differential Algorithm (SaDE) whereas the Input Selection (IS) was done using Swarm based algorithm (ACO). The simulation results revealed that a SaDE-ELM model can accurately predict *GSR* based on the satellite (Giovanni; MODIS) and observational (Reanalysis) data outperforming all other examined models. The SaDE-ELM model was tested for large scale climatic anomalies (*e.g.,* ENSO and IOD) and found that the relative errors were dramatically better than those of the comparative standalone and hybrid models (ELM, PSO-ANN, PSO-SVR etcetera). Furthermore, the sensitivity analysis using GEM-SA on predictors shows that *GSR* predictive model must incorporate the aerosol and cloud properties as a prediction variable for better accuracy. For instance, in Brisbane, when aerosol and cloud are excluded, the *RRMSE* increased to 15.818% compared to 15.980% (with only aerosol excluded) and 5.582 % (with only cloud parameters are excluded). In a similar way for Townsville, when combined aerosol and cloud parameters were excluded from the predictor variable, the *RRMSE* was 7.607% compared to 4.328% (with only aerosol excluded) and 5.543 % (with only cloud parameters excluded). Additionally, it was concluded that the predictor variables are location specific, as we have seen that there is major effect on relative error (*RRMSE* ≈15.980 %) when aerosol is excluded for Brisbane whereas for Townsville there is very minor effect (*RRMSE* ≈4.328%).

Moreover, in the third objective (Chapter 5), the swarm based IS algorithm (PSO) was used to select the most important predictors from Satellite data (MODIS). Then an advanced and non-decimated wavelet transformation known as the Maximum Overlap Discrete Wavelet Transformation (moDWT; W) was utilized in addressing non-stationarity problem whilst designing high precision *GSR* prediction model (PSO-W-SVR) for long-term basis. Three-phase PSO-W-SVR hybrid model was benchmarked with alternative methods: standalone SVR, ANN, ELM, GPML, Heteroscedatic Gaussian Processes, Adaptive Neuro-Fuzzy Inference System, Random Forest, Kernel Ridge Regression, Group Method of Data Handling and Minimax Probability Machine Regression. The Lowry plot (GEM-SA -Plot) suggest that, to consider 90% of variance, the first six parameters (Aerosol_Scattering_Angle (*asa*), Cloud_Fraction_Day (*cfd*) , Cloud_Fraction (*cf*), Atmospheric_Water_Vapor high (*awvh*), Deep_Blue_Angstrom_Exponent_Land (*dbael*) and Cirrus_Reflectance (*cr*)) are required for Alice Springs, the first four parameters (*asa, awvh,* Atmospheric_Water_Vapor low (*awvl*) and *cfd*) for Coburg and the first four parameters (*asa*, *cfd*, Cloud Top Temperature mean (*cttm*) and Cloud Top Temperature Night (*cttn*)) for Perth. Concurrent with findings from objective 2 (Chapter 4), this study also concludes that the most important input parameters are not all the same for all locations.

Furthermore, in this doctoral research thesis for fourth objective (Chapter 6) Deep Learning tools (Deep Belief Networks (DBN) and Deep Neural Networks (DNN)) with significant feature interpretation capability was designed to predict long-term *GSR*. Satellite data (MODIS) was used as predictors to predict the monthly averaged daily *GSR* as an output for four Australian solar cities. Fifteen different wrapper and filter-based IS algorithms were applied with sensitivity analysis of all MODIS-derived predictors using GEM-SA to select the optimum input for the prediction of *GSR*. The sensitivity analysis of the predictor variables demonstrated that aerosol, cloud, and water vapour parameters as input parameters play a significant role in the prediction of *GSR*. The DBN and DNN model was found to have better performances in emulating monthly averaged daily *GSR* compared to the ANN and Ensemble models (Random Forest Regression, Gradient Boosting Machine, Extreme Gradient Boosting Regression and Decision Tree).

Finally, in the fifth objective (Chapter 7), prediction of near-real-time *i.e.*, half-hourly *GSR* was achieved by designing and employing a novel model. This study has designed a Deep Learning framework, *i.e.*, CLSTM (CNN+LSTM) that integrates Convolutional Neural Networks (CNN) for pattern recognition with the Long Short-Term Memory Networks (LSTM) to construct a low latency, hybrid model. The Bureau of Meteorology (BOM) minute level *GSR* data were used. The model has been evaluated through the predictions of half-hourly, daily, and monthly solar radiations whose input elements are defined by antecedent lagged *GSR* data. In this objective, the CLSTM model was found to outperform the standalone CNN, LSTM, DNN, Recurrent Neural Network (RNN), Gated Recurrent unit Neural Network (GRU), Multilayer Perceptron (MLP) and Decision Tree (DT).

The Artificial Intelligence based predictive *GSR* model developed in this doctoral research thesis could be a particularly useful decision-support tool for energy utilisation in data sparse regions and could help facilitate core decisions about future sustainability of solar energy investments in metro, regional, and remote locations.

Additionally, incorporation of freely available Satellite and Reanalysis data with ANN, ELM, SVR and CLSTM model requires trivial human interventions. This has the prospects of being embedded into advanced prediction apps for portable devices such as tablets and mobile phones and to provide *GSR* prediction at required locations.

8.2 Limitations of the Current Study and Recommendations for Future Research

The following issues were found to be the limitations of this study, and hence are recommended in future independent studies:

- Integration of add-on optimizer algorithms; Sequential Minimal Optimization for SVR (SMO-SVR), Glow-worm Swarm Optimization algorithm for SVR (GSO-SVR)(Jiang *et al.*, 2016), Whale Optimization Algorithms for ANN (WOA-ANN) and Coral Reef Optimization for ELM (CRO-ELM) trained with Satellite-derived predictor variables could also provide greater insight into the performance of these *GSR* predictive models.
- Studies with other multiresolution analysis utilities to deal with non-stationary, such as Improved Complete Empirical Ensemble Mode Decomposition with Adaptive Noise (ICEEMDAN), Empirical Wavelet Transform (EWT), Wrapper Mutual Information Methodology (WMIM) and Variational Mode Decomposition (VMD) could may also provide greater insight into the performance of these *GSR* models.
- Alternative input selection algorithms modified Minimum Redundancy Maximum Relevance (mMRMR) algorithm, Joint Mutual Information Maximization input selection (JMIM) or Bootstrap Rank-ordered Conditional Mutual Information (broCMI) can further be explored.
- LSTM-based ELM can also be implemented to search for features that are local in space and time and its computational complexity is generally low so that it can lead to extensive feature extraction with a low latency output of the meteorological variable.

- CLSTM model was tested in only one location in Australia (Alice Springs), which can be extended to other cities and other nations with similar climate. Future research could also focus on the testing of CLSTM at different time scales, for example, at a better temporal resolution of 1-minute, 5-minute, or 10-minute prior to being implemented in Energy Management Systems.

In closing, this doctoral research has made novel contributions towards the practical problem of *GSR* forecasting using Artificial Intelligence based predictive models. This study is beneficial not only in Australia but also globally, to address climate change issues, devise new energy modelling technologies, and promote sustainable energy resources as per the United Nations Development Program Goal 7. The utilization of the free Satellite and Reanalysis dataset may be especially useful for modelling solar energy in remote regions where an installation and maintenance of any sort of ground-based equipment are not economically viable.

The findings ascertain that with appropriate input selection methods (such as the *fsrnca*, PSO or ACO) and the suitable decomposition of inputs and target data to better reveal the data features (such as using moDWT procedure), the Artificial Intelligence based predictive models can indeed capture the nonlinear dynamics and interactions among inputs and target (*GSR*) to generate an optimal model.

Moreover, the proposed Artificial Intelligence based predictive model could also help in solar plant design and could be applied to other areas such as wind speed, river flow, and electricity demand forecasting that will assist policymakers in Australia in optimal management of resources.

References

Note that the references presented here do not include the references from the published articles (Chapters 3 ,4 and 6) and the submitted manuscript (Chapter 5 and 7). These references are provided in the reference sections of the respective articles.

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). *Tensorflow: a system for large-scale machine learning.* In OSDI

Akbari, E., Buntat, Z., Enzevaee, A., Ebrahimi, M., Yazdavar, A. H., & Yusof, R. (2014). Analytical modeling and simulation of I–V characteristics in carbon nanotube based gas sensors using ANN and SVR methods. *Chemometrics and Intelligent Laboratory Systems, 137*, 173-180. doi: https://doi.org/10.1016/j.chemolab.2014.07.001. Retrieved from http://www.sciencedirect.com/science/article/pii/S0169743914001488

Akhil Raj, S. T., Venkat Ratnam, M., Narayana Rao, D., & Krishna Murthy, B. V. (2015). Vertical distribution of ozone over a tropical station: Seasonal variation and comparison with satellite (MLS, SABER) and ERA-Interim products. *Atmospheric Environment, 116*, 281-292. doi: https://doi.org/10.1016/j.atmosenv.2015.06.047. Retrieved from http://www.sciencedirect.com/science/article/pii/S1352231015301874

Alsina, E. F., Bortolini, M., Gamberi, M., & Regattieri, A. (2016). Artificial neural network optimisation for monthly average daily global solar radiation prediction. *Energy Conversion and Management, 120*, 320-329.

Antonopoulos, V. Z., Papamichail, D. M., Aschonitis, V. G., & Antonopoulos, A. V. (2019). Solar radiation estimation methods using ANN and empirical models. *Computers and Electronics in Agriculture, 160*, 160-167. doi: https://doi.org/10.1016/j.compag.2019.03.022. Retrieved from http://www.sciencedirect.com/science/article/pii/S0168169919301383

Australian Energy Update, 2018, Department of the Environment and Energy. Canberra. Retrieved from https://www.energy.gov.au/sites/default/files/australian_energy_update_2018.pdf

Bao, Y., Hu, Z., & Xiong, T. (2013). A PSO and pattern search based memetic algorithm for SVMs parameters optimization. *Neurocomputing, 117*, 98-106. doi: http://dx.doi.org/10.1016/j.neucom.2013.01.027. Retrieved from http://www.sciencedirect.com/science/article/pii/S0925231213002038

Beatley, T. (2007). Envisioning solar cities: urban futures powered by sustainable energy. *Journal of Urban Technology, 14*(2), 31-46.

Beesley, C., Frost, A., & Zajaczkowski, J. (2009). *A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets.* In 18th World IMACS/MODSIM Congress, Cairns, Australia

Belaid, S., & Mellit, A. (2016). Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate. *Energy Conversion and Management, 118*, 105-118.

Belda, M., Holtanová, E., Halenka, T., & Kalvová, J. (2014). Climate classification revisited: from Köppen to Trewartha. *Climate Research, 59*(1), 1-13.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., et al. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software, 3*(30), 774.

Bhardwaj, S., Sharma, V., Srivastava, S., Sastry, O., Bandyopadhyay, B., Chandel, S., et al. (2013). Estimation of solar radiation using a combination of Hidden Markov Model and generalized Fuzzy model. *Solar Energy, 93*, 43-54.

Birol, F. (2017). *Key World Energy Statistics 2017. International Energy Agency (IEA), 2017*. Retrieved rom URL

Blanchard, J., Martin, C., & Liu, W. (2018). Effect of ELMS and disruptions on FNSF plasma-facing components. *Fusion Engineering and Design, 135*, 337-345. doi: https://doi.org/10.1016/j.fusengdes.2017.07.022. Retrieved from http://www.sciencedirect.com/science/article/pii/S0920379617307627

BOM. (2019). *One Minute Solar Data*. Australian Bureau of Meteorology, 2019

Bouzgou, H., & Gueymard, C. A. (2017). Minimum redundancy – Maximum relevance with extreme learning machines for global solar radiation forecasting: Toward an optimized dimensionality reduction for solar time series. *Solar Energy, 158*(Supplement C), 595-609. doi: https://doi.org/10.1016/j.solener.2017.10.035. Retrieved from http://www.sciencedirect.com/science/article/pii/S0038092X17309052

Byrnes, L., Brown, C., Foster, J., & Wagner, L. D. (2013). Australian renewable energy policy: Barriers and challenges. *Renewable Energy, 60*, 711-721.

Chen, J., Zeng, G.-Q., Zhou, W., Du, W., & Lu, K.-D. (2018). Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. *Energy Conversion and Management, 165*, 681-695. doi: https://doi.org/10.1016/j.enconman.2018.03.098. Retrieved from http://www.sciencedirect.com/science/article/pii/S0196890418303261

Cheng, Y., Liu, B., Li, X., Nunziata, F., Xu, Q., Ding, X., et al. (2014). Monitoring of oil spill trajectories with COSMO-SkyMed X-Band SAR images and model simulation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7*(7), 2895-2901.

Cheng, Y., Xu, C., Mashima, D., Thing, V. L., & Wu, Y. (2017). *PowerLSTM: Power Demand Forecasting Using Long Short-Term Memory Neural Network.* In International Conference on Advanced Data Mining and Applications

Conti, J., Holtberg, P., Diefenderfer, J., LaRose, A., Turnure, J. T., & Westfall, L. 2018, *International energy outlook 2018 with projections to 2040*, USDOE Energy Information Administration (EIA), Washington, DC (United States ….

Council, C. E. (2018). *Clean energy Australia report*: Melbourne, Australia. Retrieved rom URL

Dayal, K., Deo, R., & Apan, A. A. (2017). Drought modelling based on artificial intelligence and neural network algorithms: a case study in Queensland, Australia *Climate Change Adaptation in Pacific Countries.* (pp. 177-198): Springer.

Dedekorkut, A., Mustelin, J., Howes, M., & Byrne, J. (2010). Tempering growth: planning for the challenges of climate change and growth management in SEQ. *Australian Planner, 47*(3), 203-215.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA‐Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society, 137*(656), 553-597.

Dee, D. P., & Uppala, S. (2009). Variational bias correction of satellite radiance data in the ERA‐Interim reanalysis. *Quarterly Journal of the Royal Meteorological Society, 135*(644), 1830-1841.

Deo, R. C., & Şahin, M. (2017). Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. *Renewable and Sustainable Energy Reviews, 72*, 828-848. doi: http://dx.doi.org/10.1016/j.rser.2017.01.114. Retrieved from http://www.sciencedirect.com/science/article/pii/S1364032117301247

Deo, R. C., Şahin, M., Adamowski, J. F., & Mi, J. (2019). Universally deployable extreme learning machines integrated with remotely sensed MODIS satellite predictors over Australia to forecast global solar radiation: A new approach. *Renewable and Sustainable Energy Reviews, 104*, 235-261. doi: https://doi.org/10.1016/j.rser.2019.01.009. Retrieved from http://www.sciencedirect.com/science/article/pii/S1364032119300048

Deo, R. C., Wen, X., & Qi, F. (2016). A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Applied Energy, 168*(Supplement C), 568-593. doi: https://doi.org/10.1016/j.apenergy.2016.01.130. Retrieved from http://www.sciencedirect.com/science/article/pii/S0306261916301180

Dhiman, H. S., Deb, D., & Guerrero, J. M. (2019). Hybrid machine intelligent SVR variants for wind forecasting and ramp events. *Renewable and Sustainable Energy Reviews, 108*, 369-379. doi: https://doi.org/10.1016/j.rser.2019.04.002. Retrieved from http://www.sciencedirect.com/science/article/pii/S1364032119302059

Díaz-Gómez, J., Parrales, A., Álvarez, A., Silva-Martínez, S., Colorado, D., & Hernández, J. (2015). Prediction of global solar radiation by artificial neural network based on a meteorological environmental data. *Desalination and Water Treatment, 55*(12), 3210-3217.

Dong, Z., Yang, D., Reindl, T., & Walsh, W. M. (2015). A novel hybrid approach based on self-organizing maps, support vector regression and particle swarm optimization to forecast solar irradiance. *Energy, 82*, 570-577.

Fentis, A., Bahatti, L., Mestari, M., & Chouri, B. (2017). *Short-term solar power forecasting using Support Vector Regression and feed-forward NN.* In 2017 15th IEEE International New Circuits and Systems Conference (NEWCAS)

Fisher, M. (2003). *Background error covariance modelling.* In Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean

Gala, Y., Fernández, Á., Díaz, J., & Dorronsoro, J. R. (2016). Hybrid machine learning forecasting of solar radiation values. *Neurocomputing, 176*, 48-59.

Gers, F. A., & Schmidhuber, J. (2000). *Recurrent nets that time and count.* In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium

Ghasemi, F., Mehridehnavi, A., Fassihi, A., & Pérez-Sánchez, H. (2018). Deep neural network in QSAR studies using deep belief network. *Applied Soft Computing, 62*, 251-258. doi: https://doi.org/10.1016/j.asoc.2017.09.040. Retrieved from http://www.sciencedirect.com/science/article/pii/S1568494617305793

Ghimire, S., Deo, R. C., Downs, N. J., & Raj, N. (2018). Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed MODIS satellite and Reanalysis atmospheric products in solar-rich cities. *Remote Sensing of Environment, 212*, 176-198. doi: https://doi.org/10.1016/j.rse.2018.05.003. Retrieved from http://www.sciencedirect.com/science/article/pii/S0034425718302165

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology, 377*(1-2), 80-91.

Havas, L., Ballweg, J., Penna, C., & Race, D. (2015). Power to change: Analysis of household participation in a renewable energy and energy efficiency programme in Central Australia. *Energy Policy, 87*(Supplement C), 325-333. doi: https://doi.org/10.1016/j.enpol.2015.09.017. Retrieved from http://www.sciencedirect.com/science/article/pii/S030142151530104X

Helfer, F., Lemckert, C., & Zhang, H. (2012). Impacts of climate change on temperature and evaporation from a large reservoir in Australia. *Journal of Hydrology, 475*, 365-378.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735-1780.

Hussain, S., & Al-Alili, A. (2016). A new approach for model validation in solar radiation using wavelet, phase and frequency coherence analysis. *Applied Energy, 164*, 639-649. doi: http://dx.doi.org/10.1016/j.apenergy.2015.12.038. Retrieved from http://www.sciencedirect.com/science/article/pii/S0306261915016116

Ibrahim, I. A., & Khatib, T. (2017). A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. *Energy Conversion and Management, 138*, 413-425.

IEA. (2019, 22/03/2019). *World Energy Outlook 2017*. Retrieved 22/03/2019, 2019

Jiang, H., & Dong, Y. (2016). A nonlinear support vector machine model with hard penalty function based on glowworm swarm optimization for forecasting daily global solar radiation. *Energy Conversion and Management, 126*, 991-1002. doi: https://doi.org/10.1016/j.enconman.2016.08.069. Retrieved from http://www.sciencedirect.com/science/article/pii/S0196890416307439

Kennedy, M. (2005). *GEM-SA, Version 1.1. Software: Gaussian Emulation Machine for Sensitivity Analysis*. Retrieved rom URL

Kennedy, M. C., & Petropoulos, G. P. (2017). Chapter 17 - GEM-SA: The Gaussian Emulation Machine for Sensitivity Analysis *Sensitivity Analysis in Earth Observation Modelling.* (pp. 341-361): Elsevier.

Kent, A., & Mercer, D. (2006). Australia's mandatory renewable energy target (MRET): an assessment. *Energy Policy, 34*(9), 1046-1062.

Ketkar, N. (2017). Introduction to keras *Deep Learning with Python.* (pp. 97-111): Springer.

Khatib, T., Mohamed, A., Sopian, K., & Mahmoud, M. (2012). Solar energy prediction for Malaysia using artificial neural networks. *International Journal of Photoenergy, 2012*.

Kim, H.-Y., & Liang, S. (2010). Development of a hybrid method for estimating land surface shortwave net radiation from MODIS data. *Remote Sensing of Environment, 114*(11), 2393-2402.

Kuwahata, R., & Monroy, C. R. (2011). Market stimulation of renewable-based power generation in Australia. *Renewable and Sustainable Energy Reviews, 15*(1), 534-543.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436.

Linacre, E., & Geerts, B. (1997). *Climates and weather explained*. Routledge.

Liu, H., Mi, X., & Li, Y. (2018). Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. *Energy Conversion and Management, 159*, 54-64.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing, 234*, 11-26. doi: https://doi.org/10.1016/j.neucom.2016.12.038. Retrieved from http://www.sciencedirect.com/science/article/pii/S0925231216315533

Lohmann, U., Sausen, R., Bengtsson, L., Cubasch, U., Perlwitz, J., & Roeckner, E. (1993). The Köppen climate classification as a diagnostic tool for general circulation models. *Climate Research, 3*, 177-193.

López, G., & Batlles, F. J. (2014). Estimating Solar Radiation from MODIS Data. *Energy Procedia, 49*, 2362-2369. doi: https://doi.org/10.1016/j.egypro.2014.03.250. Retrieved from http://www.sciencedirect.com/science/article/pii/S1876610214007048

Lunsford, R., Bortolon, A., Maingi, R., Mansfield, D. K., Nagy, A., Jackson, G. L., et al. (2019). Supplemental ELM control in ITER through beryllium granule injection. *Nuclear Materials and Energy, 19*, 34-41. doi: https://doi.org/10.1016/j.nme.2019.02.005. Retrieved from http://www.sciencedirect.com/science/article/pii/S235217911830125X

Mantyka‐Pringle, C. S., Martin, T. G., Moffatt, D. B., Linke, S., & Rhodes, J. R. (2014). Understanding and predicting the combined effects of climate change and land‐use change on freshwater macroinvertebrates and fish. *Journal of Applied Ecology, 51*(3), 572-581.

Marzo, A., Trigo-Gonzalez, M., Alonso-Montesinos, J., Martínez-Durbán, M., López, G., Ferrada, P., et al. (2017). Daily global solar radiation estimation in desert areas using daily extreme temperatures and extraterrestrial radiation. *Renewable Energy, 113*, 303-311. doi: http://dx.doi.org/10.1016/j.renene.2017.01.061. Retrieved from http://www.sciencedirect.com/science/article/pii/S0960148117300721

MathWorks, I. (1996). *MATLAB : the language of technical computing : computation, visualization, programming : installation guide for UNIX version 5*. Natwick : Math Works Inc., 1996. Retrieved from https://search.library.wisc.edu/catalog/9910122586102121

Mercer, P. (2014). Is Australia falling out of love with solar power. *BBC News*.

Mohammadi, K., Shamshirband, S., Tong, C. W., Arif, M., Petković, D., & Ch, S. (2015). A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. *Energy Conversion and Management, 92*, 162-171.

Mohandes, M. A. (2012). Modeling global solar radiation using Particle Swarm Optimization (PSO). *Solar Energy, 86*(11), 3137-3145. doi: https://doi.org/10.1016/j.solener.2012.08.005. Retrieved from http://www.sciencedirect.com/science/article/pii/S0038092X12002952

O'Hagan, A. (2006). Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering & System Safety, 91*(10), 1290-1300.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*(Oct), 2825-2830.

Qi, C., Zhou, Z., Sun, Y., Song, H., Hu, L., & Wang, Q. (2017). Feature selection and multiple kernel boosting framework based on PSO with mutation mechanism for hyperspectral classification. *Neurocomputing, 220*(Supplement C), 181-190. doi: https://doi.org/10.1016/j.neucom.2016.05.103. Retrieved from http://www.sciencedirect.com/science/article/pii/S0925231216309158

Qin, J., Chen, Z., Yang, K., Liang, S., & Tang, W. (2011). Estimation of monthly-mean daily global solar radiation based on MODIS and TRMM products. *Applied Energy, 88*(7), 2480-2489.

Quej, V. H., Almorox, J., Ibrakhimov, M., & Saito, L. (2017). Estimating daily global solar radiation by day of the year in six cities located in the Yucatán Peninsula, Mexico. *Journal of Cleaner Production, 141*, 75-82.

Quesada-Ruiz, S., Linares-Rodríguez, A., Ruiz-Arias, J., Pozo-Vázquez, D., & Tovar-Pescador, J. (2015). An advanced ANN-based method to estimate hourly solar radiation from multi-spectral MSG imagery. *Solar Energy, 115*, 494-504.

Ramedani, Z., Omid, M., & Keyhani, A. (2013). Modeling solar energy potential in a Tehran Province using artificial neural networks. *International Journal of Green Energy, 10*(4), 427-441.

Ramedani, Z., Omid, M., Keyhani, A., Shamshirband, S., & Khoshnevisan, B. (2014). Potential of radial basis function based support vector regression for global solar radiation prediction. *Renewable and Sustainable Energy Reviews, 39*, 1005-1011.

Rathinasamy, M., Adamowski, J., & Khosa, R. (2013). Multiscale streamflow forecasting using a new Bayesian model average based ensemble multi-wavelet Volterra nonlinear method. *Journal of Hydrology, 507*, 186-200.

Reis, A. R., Catalão, J., Vieira, G., & Nico, G. (2015). *Mitigation of atmospheric phase delay in InSAR time series using ERA-interim model, GPS and MODIS data: Application to the permafrost deformation in Hurd Peninsula, Antarctica.* In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)

REN21. (2018). *Renewables Global Status ReportRead more at: http://www.ren21.net/status-of-renewables/global-status-report/*. Renewable Energy Policy Network for 21st Century, Retrieved 17-May-2019, 2019

Renaud, O., Murtagh, F., & Starck, J.-L. (2002). *Wavelet-based forecasting of short and long memory time series*. Université de Genève/Faculté des sciences économiques et sociales.

Ryan, B. F. (1994). *Minitab handbook*. Duxbury Resource Center.

Sainath, T. N., Vinyals, O., Senior, A., & Sak, H. (2015). *Convolutional, long short-term memory, fully connected deep neural networks.* In Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on

Salcedo-Sanz, S., Deo, R. C., Cornejo-Bueno, L., Camacho-Gómez, C., & Ghimire, S. (2018). An efficient neuro-evolutionary hybrid modelling mechanism for the estimation of daily global solar radiation in the Sunshine State of Australia. *Applied Energy, 209*(Supplement C), 79-94. doi: https://doi.org/10.1016/j.apenergy.2017.10.076. Retrieved from http://www.sciencedirect.com/science/article/pii/S0306261917314976

Salcedo-Sanz, S., Deo, R. C., Cornejo-Bueno, L., Camacho-Gómez, C., & Ghimire, S. (2018). An efficient neuro-evolutionary hybrid modelling mechanism for the estimation of daily global solar radiation in the Sunshine State of Australia. *Applied Energy, 209*, 79-94.

Sanner, M. F. (1999). Python: a programming language for software integration and development. *J Mol Graph Model, 17*(1), 57-61.

Sarikprueck, P., Lee, W. J., Kulvanitchaiyanunt, A., Chen, V., & Rosenberger, J. (2017). Bounds for Optimal Control of a Regional Plug-In Electric Vehicle Charging Station System. *IEEE Transactions on Industry Applications, PP*(99), 1-1. doi: 10.1109/TIA.2017.2766230

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks, 61*, 85-117.

Şenkal, O. (2015). Solar radiation and precipitable water modeling for Turkey using artificial neural networks. *Meteorology and Atmospheric Physics, 127*(4), 481-488.

Simmons, A., Willett, K., Jones, P., Thorne, P., & Dee, D. (2010). Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets. *Journal of Geophysical Research: Atmospheres, 115*(D1).

Song, G., & Dai, Q. (2017). A novel double deep ELMs ensemble system for time series forecasting. *Knowledge-Based Systems, 134*, 31-49. doi: https://doi.org/10.1016/j.knosys.2017.07.014. Retrieved from http://www.sciencedirect.com/science/article/pii/S0950705117303295

Soufi, Y., Bechouat, M., & Kahla, S. (2017). Fuzzy-PSO controller design for maximum power point tracking in photovoltaic system. *International Journal of Hydrogen Energy, 42*(13), 8680-8688. doi: https://doi.org/10.1016/j.ijhydene.2016.07.212. Retrieved from http://www.sciencedirect.com/science/article/pii/S0360319916314501

Sözen, A., Arcaklioğlu, E., & Özalp, M. (2004). Estimation of solar potential in Turkey by artificial neural networks using meteorological and geographical data. *Energy Conversion and Management, 45*(18), 3033-3052.

Spolaore, M., Kovařík, K., Stöckel, J., Adamek, J., Dejarnac, R., Ďuran, I., et al. (2017). Electromagnetic ELM and inter-ELM filaments detected in the COMPASS Scrape-Off Layer. *Nuclear Materials and Energy, 12*, 844-851. doi: https://doi.org/10.1016/j.nme.2016.12.014. Retrieved from http://www.sciencedirect.com/science/article/pii/S2352179116301934

SunWiz. (2019). *Australian solar industry celebrates the New Year by ticking over 1.5m PV systems and one solar panel per person*. SunWiz, Retrieved 30-04-2019, 2019

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). *Sequence to sequence learning with neural networks.* In Advances in neural information processing systems

Tayal, D., & Rauland, V. (2017). Future business models for Western Australian electricity utilities. *Sustainable Energy Technologies and Assessments, 19*(Supplement C), 59-69. doi: https://doi.org/10.1016/j.seta.2016.11.007. Retrieved from http://www.sciencedirect.com/science/article/pii/S2213138816302107

The Climate Institute. (2017). *Climate of the Nation 2017 Australian attitudes on climate change*. Retrieved 13 Aug, 2017 from http://www.climateinstitute.org.au/verve/_resources/TCI0004_COTN_2017_final_version.pdf

UN, G. A. (2015). *Resolution adopted by the General Assembly on 25 September 2015. 70/1. Transforming our world: the 2030 Agenda for Sustainable Development*: New York: UN. Retrieved rom URL

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., et al. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy, 105*(Supplement C), 569-582. doi:

https://doi.org/10.1016/j.renene.2016.12.095. Retrieved from http://www.sciencedirect.com/science/article/pii/S0960148116311648

Wang, L., Kisi, O., Zounemat-Kermani, M., Zhu, Z., Gong, W., Niu, Z., et al. (2017). Prediction of solar radiation in China using different adaptive neuro-fuzzy methods and M5 model tree. *International Journal of Climatology, 37*(3), 1141-1155.

Wang, Y., Fang, Z., & Hong, H. (2019). Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. *Science of The Total Environment, 666*, 975-993. doi: https://doi.org/10.1016/j.scitotenv.2019.02.263. Retrieved from http://www.sciencedirect.com/science/article/pii/S0048969719307612

Wu, J., Chan, C. K., Zhang, Y., Xiong, B. Y., & Zhang, Q. H. (2014). Prediction of solar radiation with genetic approach combing multi-model framework. *Renewable Energy, 66*, 132-139. doi: https://doi.org/10.1016/j.renene.2013.11.064. Retrieved from http://www.sciencedirect.com/science/article/pii/S0960148113006526

Xiaoyun, Q., Xiaoning, K., Chao, Z., Shuai, J., & Xiuda, M. (2016). *Short-term prediction of wind power based on deep long short-term memory.* In 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)

You, Q., Liu, J., & Pepin, N. (2019). Changes of summer cloud water content in China from ERA-Interim reanalysis. *Global and Planetary Change, 175*, 201-210. doi: https://doi.org/10.1016/j.gloplacha.2019.02.014. Retrieved from http://www.sciencedirect.com/science/article/pii/S0921818118304302

Zahedi, A. (2016). *Solar PV for Australian tropical region; the most affordable and an appropriate power supply option.* In 2016 Australasian Universities Power Engineering Conference (AUPEC)

Zajaczkowski, J. (2009). *A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets*.

Zajaczkowski, J., Wong, K., & Carter, J. (2013). Improved historical solar radiation gridded data for Australia. *Environmental Modelling & Software, 49*, 64-77. doi: https://doi.org/10.1016/j.envsoft.2013.06.013. Retrieved from http://www.sciencedirect.com/science/article/pii/S1364815213001588

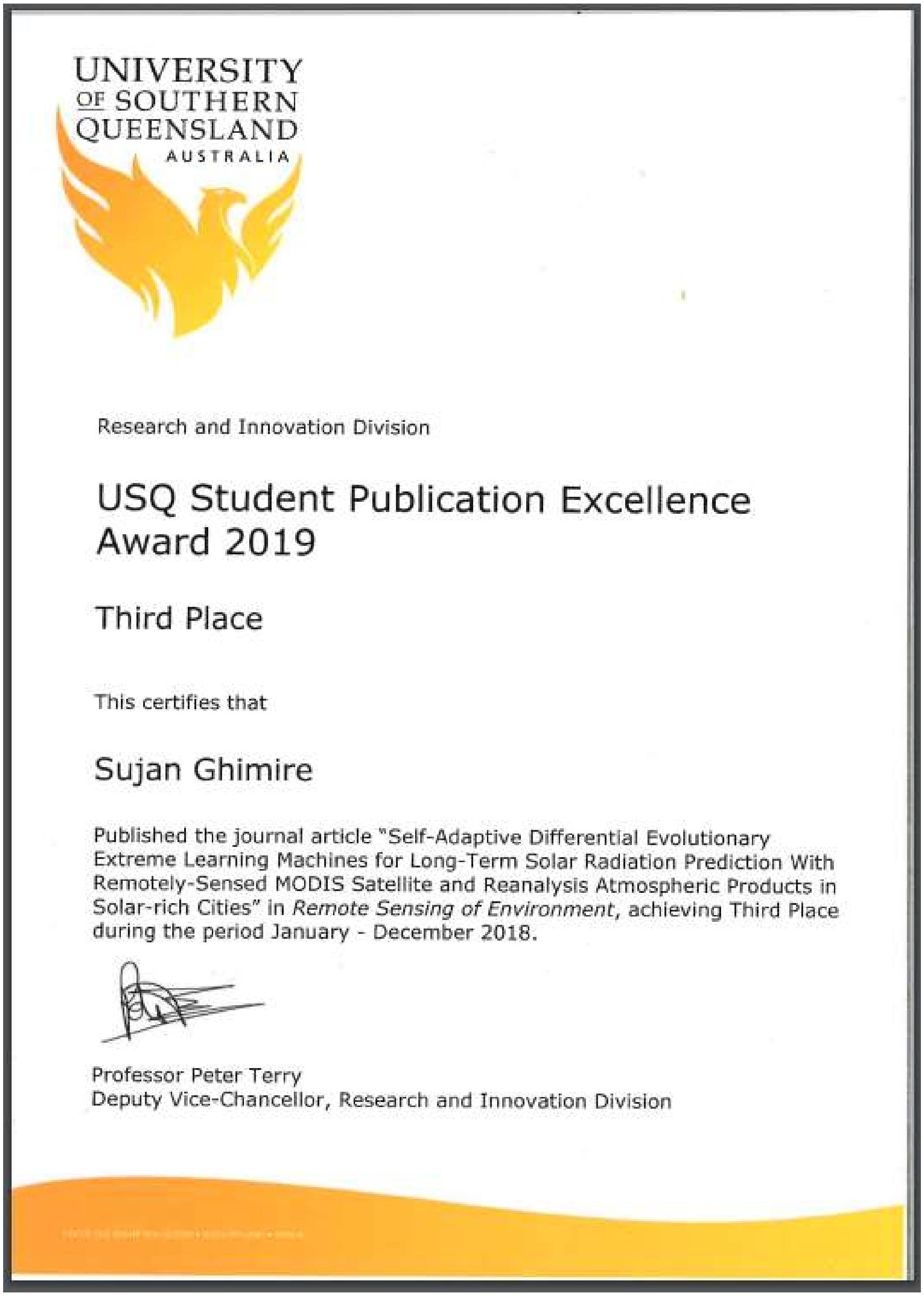
Zeng, J., & Qiao, W. (2013). Short-term solar power prediction using a support vector machine. *Renewable Energy, 52*, 118-127.

Zhou, W., Shi, J., Wang, T., Peng, B., Zhao, R., & Yu, Y. (2019). Clear-Sky Longwave Downward Radiation Estimation by Integrating MODIS Data and Ground-Based Measurements. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12*(2), 450-459.

Zhu, L., Wang, Y., & Fan, Q. (2014). MODWT-ARMA model for time series prediction. *Applied Mathematical Modelling, 38*(5), 1859-1865. doi: https://doi.org/10.1016/j.apm.2013.10.002. Retrieved from http://www.sciencedirect.com/science/article/pii/S0307904X13006148

Zhu, Y., Zhou, S., Feng, Y., Hu, Z., & Yuan, L. (2017). Influences of solar energy on the energy efficiency design index for new building ships. *International Journal of Hydrogen Energy, 42*(30), 19389-19394. doi: https://doi.org/10.1016/j.ijhydene.2017.06.042. Retrieved from http://www.sciencedirect.com/science/article/pii/S0360319917323091

Appendix 1 USQ Student publication Excellence Award 2019



UNIVERSITY OF SOUTHERN QUEENSLAND AUSTRALIA

Research and Innovation Division

USQ Student Publication Excellence Award 2019

Third Place

This certifies that

Sujan Ghimire

Published the journal article "Self-Adaptive Differential Evolutionary Extreme Learning Machines for Long-Term Solar Radiation Prediction With Remotely-Sensed MODIS Satellite and Reanalysis Atmospheric Products in Solar-rich Cities" in *Remote Sensing of Environment*, achieving Third Place during the period January - December 2018.

Professor Peter Terry
Deputy Vice-Chancellor, Research and Innovation Division

Appendix 2 Book Chapter: Optimization of Windspeed Prediction Using an Artificial Neural Network Compared With a Genetic Programming Model